Nonlinear unmixing of hyperspectral images

Yoann Altmann

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Institut National Polytechnique de Toulouse (INP Toulouse)*

Présentée et soutenue le *07/10/2013* **par :**

YOANN ALTMANN

Nonlinear spectral unmixing of hyperspectral images
Démélange non-linéaire d'images hyperspectrales

JURY

OLIVIER CAPPÉ	Directeur de recherche CNRS	Président
JOCELYN CHANUSSOT	Professeur à l'INPG/GIPSA-Lab	Rapporteur
JÉROME IDIER	Directeur de recherche CNRS	Rapporteur
CÉDRIC RICHARD	Professeur à l'Université Nice Sophia-Antipolis	Rapporteur
STEVE MCLAUGHLIN	Professeur à l'Université Heriot-Watt	Examinateur
VÉRONIQUE SERFATY	Responsable Domaine Scientifique DGA	Examinatrice
JACQUES BLANC-TALON	Responsable Domaine Scientifique DGA	Invité
MANUEL GRIZONNET	Ingénieur CNES	Invité
JEAN-YVES TOURNERET	Professeur à l'INPT-ENSEEIHT	Directeur de these
NICOLAS DOBIGEON	Maître de conférence à l'INPT-ENSEEIHT	Co-directeur de these

---

École doctorale et spécialité :
MITT : Image, Information, Hypermedia
Unité de Recherche :
Institut de Recherche en Informatique de Toulouse (UMR 5505)
Directeur(s) de Thèse :
Jean-Yves TOURNERET et *Nicolas DOBIGEON*
Rapporteurs :
Jocelyn CHANUSSOT, *Jérome IDIER* et *Cédric RICHARD*

Remerciements

Je remercie tout d'abord mon directeur de thèse Jean-Yves Tourneret et mon co-directeur de thèse Nicolas Dobigeon sans qui le travail réalisé durant ces trois ans n'aurait pas pu aboutir. Je les remercie pour leur grande disponibilité, leurs larges compétences, leurs idées et leurs commentaires qui m'ont tiré vers le haut. Je les remercie également de m'avoir donné l'opportunité de rencontrer les nombreuses personnes avec qui j'ai eu l'occasion de discuter durant cette thèse.

Je tiens à exprimer mes remerciements les plus sincères aux membres du jury de thèse qui ont accepté de juger ce travail. Je suis profondément reconnaissant à mes trois rapporteurs, M. Jocelyn Chanussot, M. Jérôme Idier et M. Cédric Richard, d'avoir accepté de lire et évaluer ma thèse. Je remercie également M. Olivier Cappé, M. Steve McLaughlin, et Mme Véronique Serfaty, qui ont accepté d'être examinateurs de ma thèse.

Je remercie tous mes collègues de l'IRIT, enseignants, chercheurs, doctorants, techniciens, secrétaires, pour leur sympathie et convivialité au sein du laboratoire. Merci pour les aides permanentes reçues du personnel du laboratoire.

Je n'oublie pas mes visites à Edimbourg durant lesquelles j'ai eu l'occasion de rencontrer et de discuter avec de nombreuses personnes et je remercie M. Steve McLaughlin pour son accueil, sa convivialité et sa disponibilité. Je suis persuadé que le post-doc à Heriot-Watt sera une expérience exceptionnelle.

Je remercie particulièrement Abderrahim, Cécile, Julien Renaud, Bouchra, Soksok, Olivier, Romain, de l'IRIT, sans oublier Raoul, Victor, Florian, Chao, Jorge et Jean-Adrien du Tésa pour ces trois années.

Une pensée à ma famille et mes amis qui m'ont encouragé durant ma thèse. Je remercie mes parents de m'avoir permis de poursuivre mes études et je leur exprime toute mon affection et ma gratitude.

Enfin, et j'aurais tout aussi bien pu commencer ces remerciements par elle, je remercie *amore mio*, celle qui m'accompagne, me soutient et qui a largement contribué à cette thèse. Je lui doit beaucoup et plus encore.

Merci à tous.

Résumé

Le démélange spectral est un des sujets majeurs de l'analyse d'images hyperspectrales. Ce problème consiste à identifier les composants macroscopiques présents dans une image hyperspectrale et à quantifier les proportions (ou abondances) de ces matériaux dans tous les pixels de l'image. La plupart des algorithmes de démélange suppose un modèle de mélange linéaire qui est souvent considéré comme une approximation au premier ordre du mélange réel. Cependant, le modèle linéaire peut ne pas être adapté pour certaines images associées par exemple à des scènes engendrant des trajets multiples (forêts, zones urbaines) et des modèles non-linéaires plus complexes doivent alors être utilisés pour analyser de telles images.

Le but de cette thèse est d'étudier de nouveaux modèles de mélange non-linéaires et de proposer des algorithmes associés pour l'analyse d'images hyperspectrales. Dans un premier temps, un modèle paramétrique post-non-linéaire est étudié et des algorithmes d'estimation basés sur ce modèle sont proposés. Les connaissances a priori disponibles sur les signatures spectrales des composants purs, sur les abondances et les paramètres de la non-linéarité sont exploitées à l'aide d'une approche bayesienne. Le second modèle étudié dans cette thèse est basé sur l'approximation de la variété non-linéaire contenant les données observées à l'aide de processus gaussiens. L'algorithme de démélange associé permet d'estimer la relation non-linéaire entre les abondances des matériaux et les pixels observés sans introduire explicitement les signatures spectrales des composants dans le modèle de mélange. Ces signatures spectrales sont estimées dans un second temps par prédiction à base de processus gaussiens.

La prise en compte d'effets non-linéaires dans les images hyperspectrales nécessite souvent des stratégies de démélange plus complexes que celles basées sur un modèle linéaire. Comme le modèle linéaire est souvent suffisant pour approcher la plupart des mélanges réels, il est intéressant de pouvoir détecter les pixels ou les régions de l'image où ce modèle linéaire est approprié. On pourra alors, après cette détection, appliquer les algorithmes de démélange non-linéaires aux pixels nécessitant réellement l'utilisation de modèles de mélange non-linéaires. La dernière partie de ce manuscrit se concentre sur l'étude de détecteurs de non-linéarités basés sur des modèles linéaires et non-linéaires pour l'analyse d'images hyperspectrales.

Les méthodes de démélange non-linéaires proposées permettent d'améliorer la caractérisation des images hyperspectrales par rapport au méthodes basées sur un modèle linéaire. Cette amélioration se traduit en particulier par une meilleure erreur de reconstruction des données. De plus, ces méthodes permettent de meilleures estimations des signatures spectrales et des abondances quand les pixels résultent de mélanges non-linéaires. Les résultats de simulations effectuées sur des données synthétiques et réelles montrent l'intérêt d'utiliser des méthodes de détection de non-linéarités pour l'analyse d'images hyperspectrales. En particulier, ces détecteurs peuvent permettre d'identifier des composants très peu représentés et de localiser des régions où les effets non-linéaires sont non-négligeables (ombres, reliefs,...). Enfin, la considération de corrélations spatiales dans les images hyperspectrales peut améliorer les performances des algorithmes de démélange non-linéaires et des détecteurs de non-linéarités.

Abstract

Spectral unmixing is one the major issues arising when analyzing hyperspectral images. It consists of identifying the macroscopic materials present in a hyperspectral image and quantifying the proportions of these materials in the image pixels. Most unmixing techniques rely on a linear mixing model which is often considered as a first approximation of the actual mixtures. However, the linear model can be inaccurate for some specific images (for instance images of scenes involving multiple reflections) and more complex nonlinear models must then be considered to analyze such images.

The aim of this thesis is to study new nonlinear mixing models and to propose associated algorithms to analyze hyperspectral images. First, a post-nonlinear model is investigated and efficient unmixing algorithms based on this model are proposed. The prior knowledge about the components present in the observed image, their proportions and the nonlinearity parameters is considered using Bayesian inference. The second model considered in this work is based on the approximation of the nonlinear manifold which contains the observed pixels using Gaussian processes. The proposed algorithm estimates the relation between the observations and the unknown material proportions without explicit dependency on the material spectral signatures, which are estimated subsequentially.

Considering nonlinear effects in hyperspectral images usually requires more complex unmixing strategies than those assuming linear mixtures. Since the linear mixing model is often sufficient to approximate accurately most actual mixtures, it is interesting to detect pixels or regions where the linear model is accurate. This nonlinearity detection can be applied as a pre-processing step and nonlinear unmixing strategies can then be applied only to pixels requiring the use of nonlinear models. The last part of this thesis focuses on new nonlinearity detectors based on linear and nonlinear models to identify pixels or regions where nonlinear effects occur in hyperspectral images.

The proposed nonlinear unmixing algorithms improve the characterization of hyperspectral images compared to methods based on a linear model. These methods allow the reconstruction errors to be reduced. Moreover, these methods provide better spectral signature and abundance estimates when the observed pixels result from nonlinear mixtures. The simulation results conducted on synthetic and real images illustrate the advantage of using nonlinearity detectors for hyperspectral image analysis. In particular, the proposed detectors can identify components which are present in few pixels (and hardly distinguishable) and locate areas where significant nonlinear effects occur (shadow, relief, ...). Moreover, it is shown that considering spatial correlation in hyperspectral images can improve the performance of nonlinear unmixing and nonlinearity detection algorithms.

Acronymes and notations

Acronymes

ARE	average reconstruction error
CCRLB	constrained Cramér-Rao lower bound
CHMC	constrained Hamiltonian Monte Carlo
CRLB	Cramér-Rao lower bound
CHMC	constrained Hamiltonian Monte Carlo
DAG	directed acyclic graph
EEA	endmember extraction algorithm
FCLL-GPLVM	fully constrained locally linear Gaussian process latent variable model
FM	Fan bilinear model
GBM	generalized bilinear model
GLRT	generalized likelihood ratio test
GP	Gaussian process
GPLVM	Gaussian process latent variable model
HMC	Hamiltonian Monte Carlo
i.i.d.	independent and identically distributed
K-FCLS	kernel fully constrained least squares
LL-GPLVM	locally linear Gaussian process latent variable model
LMM	linear mixing model
LS	least-squares
MAP	maximum a posteriori
MCMC	Markov chain Monte Carlo
MMSE	minimum mean square error
MLE	maximum likelihood estimator
MRF	Markov random field
MSE	mean square error
NLMM	nonlinear mixing model
PCA	principal component analysis
PPNMM	post nonlinear mixing model
RCA-SU	residual component analysis-based spectral unmixing
RNMSE	root normalized mean square error

RKHS	reproducing kernel Hilbert space
SAM	spectral angle mapper
SNR	signal-to-noise ratio
SO	subgradient optimization
SU	spectral unmixing

Standard notations

$\propto$	proportional to
$\sim$	distributed according to
$\ll$	much lower
$\gg$	much greater
$\Gamma(\cdot)$	Gamma function
$B(\cdot,\cdot)$	Beta function
$\delta(\cdot)$	Dirac delta function
$\odot$	Hadamard product
$\otimes$	Kronecker product
$\mathbf{1}_{\mathbb{E}}(\mathbf{x})$	indicator function defined on $\mathbb{E}$ $\begin{cases} \mathbf{1}_{\mathbb{E}}(\mathbf{x}) = 1 & \text{if } \mathbf{x} \in \mathbb{E} \\ \mathbf{1}_{\mathbb{E}}(\mathbf{x}) = 0 & \text{else.} \end{cases}$

Matrix notations

x	scalar value
$\mathbf{x}$	column vector
$.^T$	transpose operator
$\mathbf{X}$	matrix
$\mathbf{x}_n$	nth column of the matrix $\mathbf{X}$
$\mathbf{x}_{n,:}$	column vector consisting of the nth row of the matrix $\mathbf{X}$
$\mathrm{tr}(\mathbf{X})$	trace of the matrix $\mathbf{X}$
$\mathrm{etr}(\mathbf{X})$	exponential trace of the matrix $\mathbf{X}$
$\lvert\mathbf{X}\rvert$	determinant of the matrix $\mathbf{X}$
$\mathbf{1}_d$	$d \times 1$ ones vector
$\mathbf{0}_d$	$d \times 1$ zeros vector
$\mathbf{I}_d$	$d \times d$ identity matrix
$\lVert\mathbf{x}\rVert$	standard ℓ^2-norm $\lVert\mathbf{x}\rVert = \sqrt{\mathbf{x}^T\mathbf{x}}$

Spectral unmixing notations

R	number of endmembers
r	endmember index
N	number of pixels
n	pixel index
L	number of spectral bands
ℓ	band index
$\mathbf{y}$	pixel spectrum
$\boldsymbol{a}$	abundance vector
$\mathbf{m}_r$	rth endmember spectrum
$\mathbf{M}$	endmember matrix

Sampling notations

N_{MC}	length of the Markov chain
N_{bi}	length of the burn-in period of the Markov chain
N_{LF}	number of leap-frog steps
$x^{(k)}$	k-th sample of the chain $\left(x^{(k)}\right)_{k=1,\ldots,N_{\mathrm{MC}}}$

Usual distributions

$\mathcal{U}_{\mathbb{E}}(\mathbf{x})$	uniform distribution defined on the set $\mathbb{E}$
$\mathcal{N}\left(m,\sigma^2\right)$	Gaussian distribution with mean m and variance σ^2
$\mathcal{N}_{\mathbb{E}}(m,\sigma^2)$	truncated Gaussian distribution, whose support is $\mathbb{E}$ and with hidden mean m and hidden variance σ^2
$\mathcal{N}\left(\mathbf{m},\boldsymbol{\Sigma}\right)$	multivariate Gaussian distribution with mean $\mathbf{m}$ and covariance matrix $\boldsymbol{\Sigma}$
$\mathcal{N}_{\mathbb{E}}(\mathbf{m},\boldsymbol{\Sigma})$	truncated multivariate Gaussian distribution, whose support is $\mathbb{E}$ and with hidden mean $\mathbf{m}$ and hidden covariance matrix $\boldsymbol{\Sigma}$
$\mathcal{IG}(\gamma,\nu)$	inverse-gamma distribution with shape parameter γ and scale parameter ν.
$\mathcal{B}e(\alpha,\beta)$	Beta distribution with shape parameters α and β.

Contents

List of Figures

List of algorithms

List of Tables

Introduction (in French)

Contexte et problématique de la thèse

Le travail de thèse présenté dans ce manuscrit a été effectué à l'Institut de Recherche en Informatique de Toulouse (IRIT), dans l'équipe "Signal et Communication". Cette équipe possède une forte expertise dans le développement de méthodes de traitement d'images hyperspectrales et le travail de recherche réalisé pendant cette thèse s'inscrit naturellement dans la continuité des travaux effectué par Nicolas Dobigeon et Olivier Eches sur le démélange spectral linéaire. Cette thèse a été financée pour une durée de trois ans par la Direction Générale de l'Armement (DGA). Ce manuscrit présente de nouveaux modèles non-linéaires et algorithmes associés pour le démélange spectral d'images hyperspectrales. Le démélange spectral (SU pour *spectral unmixing*) est un problème de séparation de sources qui suscite un fort intérêt depuis les dernières décennies dans les domaines du traîtement du signal et des images. Il consiste à extraire de données hyperspectrales (ici d'une image), les composantes spectrales des composants macroscopiques purs contenus dans les données et appelées *endmembers* et à identifier leurs interactions appelés mélanges. Le démélange spectral, comme la plupart des problèmes de séparation de sources, est un problème difficile et mal-posé. Pour réduire la complexité du problème, la plupart des travaux de la litérature sont basés sur un modèle de mélange linéaire (LMM pour *linear mixing model*), considéré comme une première approximation du modèle de mélange réel. Ce modèle est dit linéaire dans le sens où le spectre observé d'un pixel mélangé résulte d'une combinaison linéaire des endmembers. Cependant, cette approximation peut être trop éloignée du modèle réel pour certaines images et des modèles plus complexes doivent être utilisés pour dépasser les limitations intrinsèques du LMM. Plusieurs modèles de mélanges non-linéaires (NLMMs pour *nonlinear mixing models*) pour le démélange spectral ont été proposés pour caractériser les effets non-linéaires qui peuvent apparaitre dans les images hyperspectrales. A cause de la grande diversité de ces effets, la conception et/ou le choix d'un modèle approprié au problème de démélange est difficile. Les récents travaux de la litérature se concentrent sur de nouveaux modèles non-linéaires et algorithmes de démélange associés.

Structure du manuscrit

Le premier chapitre de ce manuscrit présente un modèle post-non-linéaire pour le démélange spectral non-linéaire. Le modèle post-non-linéaire polynomial (PPNMM) étudié permet de modéliser des non-linéarités qui peuvent différer d'un pixel à l'autre. Une propriété importante de ce modèle est que la non-linéarité de chaque pixel est caractérisée par un seul paramètre. De plus, ce paramètre est nul si le modèle de mélange sous-jacent est linéaire. La première partie de ce chapitre se concentre sur le problème de démélange dans le cas où les signatures spectrales des composants purs de l'image sont connues. Le premier algorithme développé est constitué d'un modèle bayésien hiérarchique couplé à des méthodes de simulation MCMC (Markov chain Monte Carlo). Deux autres algorithmes d'optimisation sont également proposés pour résoudre le problème avec un coût calculatoire réduit par rapport à l'algorithme bayésien. La flexibilité du modèle PPNMM et les performances des trois méthodes sont analysées à

l'aide de simulations réalisées sur des données synthétiques et réelles. La seconde partie du premier chapitre étudie un nouveau modèle bayésien pour résoudre le problème de démélange à l'aide du modèle PPNMM dans le cas où les signatures spectrales sont inconnues et doivent être estimées. Des lois a priori sont associées à tous les paramètres inconnus du modèle. L'estimation conjointe des signatures spectrales et des coefficients de mélange induit un grand nombre de paramètres à simuler. Pour améliorer l'efficacité de l'échantillonneur, des méthodes de Monte Carlo à base de dynamiques hamiltoniennes contraintes sont utilisées.

Le second chapitre de ce manuscrit étudie une nouvelle méthode à base de noyaux pour le démélange spectral non-linéaire. Cette méthode est étroitement liée à une récente méthode de réduction de dimension non-linéaire utilisant un modèle à variables latentes et des processus gaussiens (*Gaussian process latent variable model* (GPLVM)) (Lawrence, 2003). Les GPLVMs, qui succitent de plus en plus d'intérêt dans le domaine de l'apprentissage automatique, ont la propriété intéressante de pouvoir approcher de nombreuses relations non-linéaires entre un espace dit "latent" (de dimension faible) et l'espace des observations (de dimension plus importante). Les relations non-linéaires et souvent complexes peuvent être approchées par des fonctions à base de noyaux (Perez-Cruz et al., 2013). Par conséquent, ces modèles sont particulièrement adaptés pour analyser des images hyperspectrales et potentiellement pour résoudre le problème de démélange. Dans ce chapitre, nous proposons d'utiliser une forme particulière de noyau basé sur les modèles bilinéaires existants, ce qui permet à l'algorithme de démélange proposé d'être bien adapté lorsque le modèle de mélange sous-jacent est bilinéaire. La première étape de l'algorithme bayésien proposé consiste à estimer les abondances des pixels de l'image. Des lois a priori choisies avec soin sont affectées aux paramètres inconnus du modèle GPLVM (y compris les variables latentes liées aux abondances) à estimer. La distribution a posteriori de ces paramètres est obtenue en utilisant l'inférence bayésienne. Les estimateurs du maximum a posteriori (MAP) des paramètres sont obtenus en utilisant une méthode de gradient et les abondances sont estimées à partir des variables latentes estimées. Une fois les abondances estimées, les signatures spectrales des composants purs de l'image sont estimées à l'aide de la capacité de prédiction des processus gaussiens. L'algorithme de démélange proposé brise le paradigme habituel du démélange spectral en estimant d'abord les abondances et en prédisant ensuite les signatures spectrales. Les simulations réalisées sur des images synthétiques et réelles illustrent la flexibilité du modèle proposé pour le démélange spectral linéaire et non linéaire et donnent des résultats prometteurs concernant l'estimation des abondances et des spectres d'intérêt, même en l'absence de pixels purs dans l'image.

Le troisième chapitre de ce manuscrit est dédié à la détection de non-linéarités dans les images hyperspectrales. Ce chapitre se concentre sur des procédures de détection supervisées, c'est-à-dire quand les signatures spectrales des composants purs de l'image sont *a priori* connues. Le premier détecteur supervisé est basé sur le modèle PPNMM étudié dans le premier chapitre. Plus précisément, le test associé est basé sur les propriétés statistiques des paramètres de ce modèle et permet de décider si un pixel donné résulte d'un mélange linéaire ou non linéaire. Inversement, le deuxième détecteur supervisé ne suppose aucun modèle non-linéaire particulier et repose uniquement sur le modèle de mélange linéaire. Les deux détecteurs sont comparés en utilisant des simulations effectuées sur des données synthétiques et réelles.

Le détecteur de non-linéarités basé sur le modèle PPNMM et étudié dans le troisième chapitre est utilisé après une procédure de délémange basée sur ce même modèle. D'un autre côté, le détecteur de non-linéarité basé sur le modèle LMM ne nécessite pas explicitement une étape de démélage. Cependant, il serait intéressant de proposer des algorithmes qui permettent d'effectuer simultanément le démélange spectral (linéaire/non-linéaire) et la détection de non-linéarités. Le dernier chapitre de ce manuscrit présente un modèle de mélange non-linéaire pour effectuer

conjointement le démélange et la détection de non-linéarités. Le modèle non-linéaire proposé suppose que les réflectances des pixels sont des combinaisons linéaires des composantes spectrales pures connus, affectées par un terme additive qui dépend de ces signatures spectrales et contaminées par un bruit additif. Un champ de Markov est considéré pour la détection de non-linéarités afin de prendre en compte la structure spatiale des termes non-linéaires. L'image observée est alors segmentée en régions où ces termes non-linéaires, s'ils sont présents, partagent les mêmes propriétés statistiques. Un algorithme bayésien est proposé pour estimer les paramètres du modèle, ce qui permet d'effectuer simultanément le démélange spectral et la détection de non-linéarités. Les performances de l'algorithme sont d'abord évaluées sur des données synthétiques et les simulations effectuées avec des données réelles donnent des résultats intéressants et prometteurs.

Contributions majeures

- **Premier chapitre.** Un nouveau modèle non-linéaire à base de mélanges post-non-linéaires est proposé pour le démélange linéaire/non-linéaire. La non-linéarité de chaque pixel est caractérisée par un paramètre d'amplitude unique, ce qui permettra de proposer des détecteurs de non-linéarités simples dans le troisième chapitre de cette thèse. Des méthodes efficaces sont proposées pour résoudre les problèmes de démélange supervisé et non-supervisé. Des algorithmes d'échantillonnage basés sur des méthodes MCMC à dynamiques hamiltoniennes sont également utilisés pour améliorer les propriétés de mélange de l'échantillonneur lorsque les composantes spectrales pures de l'image sont inconnues (et doivent être estimées).

- **Deuxième chapitre.** Un nouveau modèle de mélange non-linéaire et non-paramétrique est proposé. L'algorithme de démélange associé brise le paradigme habituel du démélange spectral en estimant d'abord les abondances et en prédisant ensuite les signatures spectrales, ce qui se révèle efficace en cas d'absence de pixels purs dans l'image.

- **Troisième chapitre.** Des tests statistiques sont proposés pour la détection de non-linéarités affectant les pixels d'une image hyperspectrale, lorsque les composants de cette image sont connus. Les tests proposés sont simples à mettre en oeuvre, avec un coût calculatoire faible pour pouvoir être appliqués en pratique. Les résultats de détection peuvent alors être utilisés comme une étape de pré-traitement pour sélectionner des pixels ou des régions où des modèles plus complexes doivent être utilisés à la place du modèle linéaire classique pour le démélange spectral.

- **Quatrième chapitre.** Un nouveau modèle de mélange non-linéaire pour effectuer conjointement le démélange d'image hyperspectrale et la détection de non-linéarités est proposé. L'image observée est segmentée en régions où les termes non-linéaires, s'ils sont présents, partagent les mêmes propriétés statistiques. L' algorithme proposé permet d'estimer convenablement les abondances lorsque les mélanges réels sont linéaires et non-linéaires et il généralise les détecteurs de non-linéarité binaires proposées dans le troisième chapitre en considérant différents niveaux (classes) de non-linéarités.

Introduction

Context and objectives of the thesis

This thesis has been achieved in the Institut de Recherche en Informatique de Toulouse (IRIT), within the "Signal and Communication group". This group has a strong experience in developing hyperspectral image processing methods. The work presented herein follows naturally previous works conducted by Nicolas Dobigeon and Olivier Eches on linear spectral unmixing. This thesis has been funded by Direction Générale de l'Armement (DGA), French Ministry of defence, for a three-year period. This manuscript presents new nonlinear models and associated algorithms for spectral unmixing of hyperspectral images. Spectral unmixing (SU) is a source separation problem that has received intensive interest over the few last decades in signal and image processing (Bioucas-Dias et al., 2012; Craig, 1994; Keshava and Mustard, 2002). It consists of extracting from a hyperspectral image, the spectra of the pure macroscopic components present in the image, referred to as endmembers, and of identifying their interactions or mixtures. SU of hyperspectral images, as most of the source separation tasks, is a challenging ill-posed problem. To reduce the problem complexity, most works of the literature rely on a linear mixing model (LMM), often considered as a first approximation of the actual mixture. This model is referred to as linear in the sense that the reflectance spectrum of a mixed pixel is assumed to be a linear combination of the endmembers. However, this approximation can be inaccurate to describe some scenes, requiring more complex mixing models to be considered (to overcome the inherent limitations of the LMM). Few studies have addressed the problem of nonlinear SU to characterize nonlinear effects affecting hyperspectral images. Designing and/or choosing an appropriate nonlinear mixing model (NLMM) for SU is a challenging problem because of the diversity of nonlinear effects. Second, introducing nonlinear terms in the observation model complicates the derivation of efficient nonlinear SU procedures relying on NLMMs. The aim of this work is to propose efficient image analysis and nonlinear SU methods based on flexible NLMMs.

Linear spectral unmixing of hyperspectral images

Hyperspectral images are 3-dimensional data cubes. These cubes consist of two spatial and one spectral dimensions. The spectral dimension corresponds to the different wavelengths at which the scene is observed. Each pixel of a multidimensional image corresponds to an elementary surface of the observed scene and is represented by a vector of values, whose length is given by the number of spectral bands considered. Classical colored images can be seen as particular multiband images where the number of spectral bands reduces to three (red, green and blue). When the number of spectral bands increases, the images are called multispectral (up to about 10 bands) and

hyperspectral images composed of several hundreds of bands. Multispectral and thus hyperspectral images offer a better spectral resolution when compared to classical RGB images. These images allow elements in a scene to be easily distinguished using more spectral information available, which could be difficult or even impossible when using other kinds of images. The hyperspectral images analyzed in this manuscript are reflectance images, i.e., the vector of values associated with each pixel is the reflectance vector (or spectrum) of the corresponding surface in the scene. Consequently, the spectrum of a pixel composed of a sole material (such as soil vegetation or water) is characteristic of this endmember (see Fig. 1). Based on this observation, many works have been conducted to segment hyperspectral images and to classify the image pixels into groups sharing similar spectral properties. This classification problem has received intensive interest and provided interesting results in terms of scene understanding (Fauvel et al., 2013; Tarabalka et al., 2010).

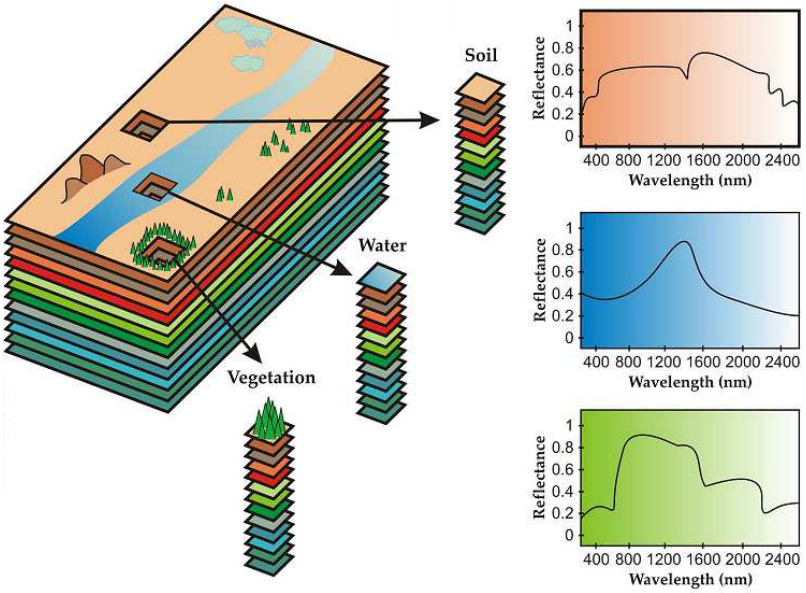


Figure 1: Example of hyperspectral imaging concept.

Current spectrometers used to acquire remote sensing hyperspectral images generally have a spatial resolution of $20m$ to $0.5m$. Due to the spatial resolution of the hyperspectral imaging systems, several components can lye within the same pixel. The resulting spectrum is thus a mixture of the spectra of each component, which makes the classification less obvious. To alleviate the limitations of a hard classification, unmixing techniques have been proposed for a more precise scene characterization which can be considered as a soft classification (Kent and Mardia, 1988). The next section recalls the classical linear mixing model used to address the problem of hyperspectral image SU.

Linear mixing model

As mentioned above, hyperspectral images have two spatial dimensions and thus the pixel positions are characterized by two spatial coordinates, i.e., a row and a column index. However, a single position index is preferred in this manuscript for ease of reading. Let N_{row} (resp. N_{col} and L) denote the number of rows (resp. columns and spectral bands) of the observed image. The image contains $N = N_{\text{row}} \times N_{\text{col}}$ pixels $\mathbf{y}_n (n = 1, \ldots, N)$ observed at L different wavelengths. When each photon reaching the sensor has interacted with a sole macroscopic component of the

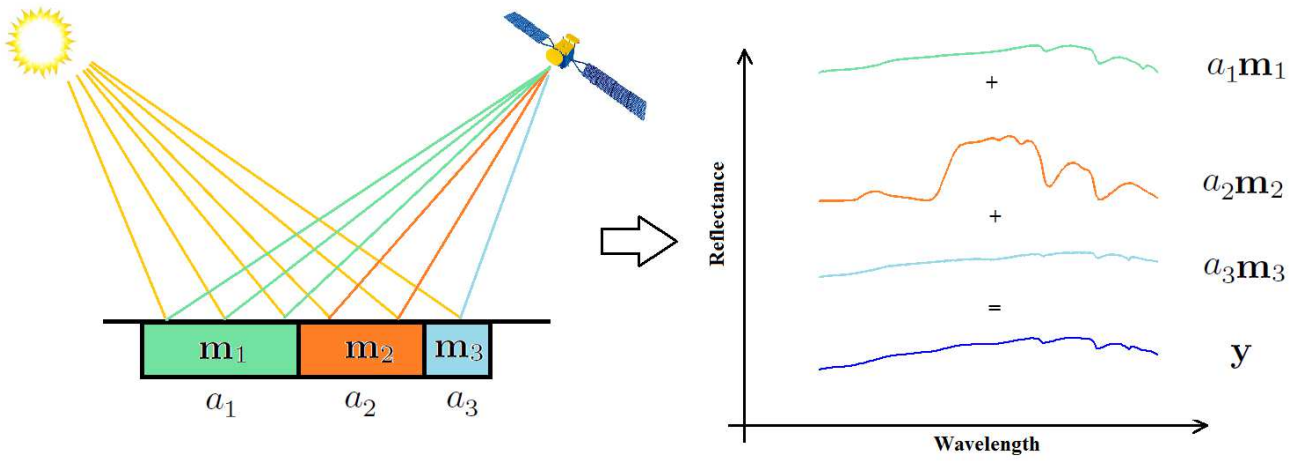


Figure 2: Linear mixing model: the area of interest is flat and is composed of components sitting side-by-side.

observed scene, the nth measured reflectance spectrum $\mathbf{y}_n \in \mathbb{R}^L$ can be accurately described by the following LMM

$$\mathbf{y}_n = \sum_{r=1}^{R} a_{r,n}\mathbf{m}_r + \mathbf{e}_n \tag{1}$$

where R is the number of endmembers in the image, $\mathbf{m}_r = [m_{r,1}, \ldots, m_{r,L}]^T$ is the spectrum of the rth endmember, $a_{r,n}$ is its corresponding proportion (or *abundance*) in the nth pixel and $\mathbf{e}_n$ is an additive noise sequence associated with the noise measurement and the model error. This photon behavior typically occurs when the observed scene is flat and when the macroscopic elements sit side-by-side in the scene, as a checker board structure (see example in Fig. 2). In this case, the abundances can be seen as relative surfaces occupied by each component in each pixel. Thus, it is reasonable to consider the following constraints

$$\begin{cases} a_{r,n} > 0, \forall r, \quad \forall n \\ \sum_{r=1}^{R} a_{r,n} = 1, \quad \forall n \end{cases} \tag{2}$$

for the abundances. These constraints are referred to as positivity and sum-to-one constraints. Most dimensionality reduction techniques and SU algorithms rely on these constraints for the abundances. However, the sum-to-one constraint is sometimes neglected. In this manuscript, both constraints are used when considering the LMM.

Unmixing strategies

SU (even when assuming the LMM) is a challenging blind source separation problem often split into two steps.

- **Endmember estimation:**
  Endmember estimation is a crucial step which is the first step for the analysis of a hyperspectral image. Many geometrical LMM-based endmember extraction algorithms (EEA) have been proposed in the literature and can be classified into two groups. The first group consists of methods looking for purest pixels in the data set. This group includes the pixel purity index (PPI) (Chaudhry et al., 2006), vertex component analysis (VCA)

(Nascimento and Bioucas-Dias, 2005) and N-FINDR (Winter, 1999) algorithms. The second group estimates sets of smallest volumes embedding the data, such as the minimum volume simplex analysis (MVSA) (Li and Bioucas-Dias, 2008) algorithm.

- **Inversion:**
  Once the endmembers have been extracted from the data or from a spectral library, the second step, referred to as inversion, estimates the abundance vectors of the image pixels. The most popular inversion algorithm is probably the fully constrained least squares (FCLS) algorithm (Heinz and C.-I Chang, 2001) which has shown excellent performance in many practical applications. Bayesian methods have also be proposed to achieve the inversion step (Dobigeon et al., 2008). The Bayesian framework is particularly well adapted to include uncertainties about the endmembers estimated during the previous step (Eches et al., 2010) and spatial correlations (Eches et al., 2011) within the abundance estimation step.

More recently, methods have been proposed for jointly estimating the pure spectral signatures and the abundances.

- **Joint endmember and abundance estimation approaches:**
  These methods generally overcome the methods decomposed into two successive steps in terms of endmember and abundance estimation. Thus, they can be more robust when they are too few pure pixels in the image (absence of endmembers) and in the presence of outliers (measurement errors,...). Joint endmember and abundance estimation can be expressed as a nonnegative matrix factorization (NMF) problem. The main works studying joint endmember and abundance estimation include the iterative constrained endmembers (ICE) (Berman et al., 2004) algorithm, the minimum volume constrained NMF (MVC-NMF) (Miao and Qi, 2007) and Bayesian algorithms (Dobigeon et al., 2009a,b)

A more detailed review on LMM-based SU procedures has been recently proposed by Bioucas-Dias et al. (2012). It is important to distinguish two kinds of unmixing strategies depending on the prior knowledge about the observed image. Supervised methods assume that the endmembers present in the image are a priori known. In this case, the SU procedures reduce to the inversion step. Conversely, unsupervised methods first estimate the endmembers (number and spectra) and then consider the inversion step. When the endmembers are partially known (number of components known for instance), the associated methods are sometimes called semi-supervised. However, these methods are considered as unsupervised in this manuscript, for clarity.

Nonlinear unmixing procedures

Due to their relative simplicity, SU procedures assuming the LMM have been widely used for the analysis of hyperspectral image. However, it has been pointed out that the LMM can be inaccurate in particular situations (Bioucas-Dias et al., 2012; Keshava and Mustard, 2002). For these cases, more complex models must be used. Several approximations based on the radiative transfer theory have been proposed such as the bidirectional model proposed by Hapke (1981). However, these approximations require highly nonlinear formulations which complicate the derivation of unmixing strategies. Alternative approximations have been proposed for handling nonlinear effects leading to exploitable physics-based nonlinear mixing models.

Bilinear mixing models

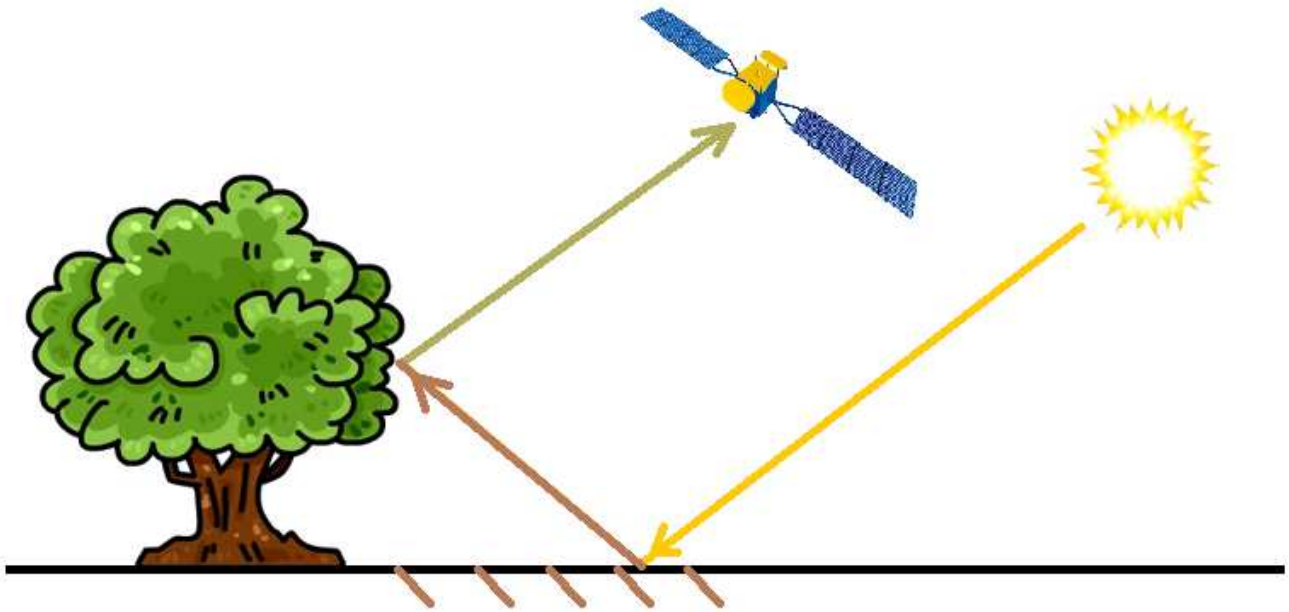


Figure 3: Nonlinear mixtures: example of multiple reflections occurring in forested areas.

Bilinear models have received growing interest over the last few years for their ability to capture multiple scattering effects in hyperspectral images. Such phenomena may occur when the light scattered by a given material reflects off other surfaces before reaching the sensor. This is typically the case when the observed scene contains forested or urban areas, where interactions occur between the ground and the canopy or buildings as illustrated in Fig. 3. Most bilinear models proposed in the literature (to address the multiple reflection problem) can be expressed as

$$\mathbf{y}_n = \sum_{r=1}^{R} a_{r,n}\mathbf{m}_r + \sum_{i=1}^{R-1}\sum_{j=i+1}^{R} \beta_{i,j,n}\mathbf{m}_i \odot \mathbf{m}_j + \mathbf{e}_n \tag{3}$$

where $\odot$ denotes the Hadamard (termwise) product. The first term in the right-hand side of (3) contains the linear part of the mixture while the double sum models the nonlinear effects. The additional parameter $\beta_{i,j,n}$ is an amplitude coefficient that tunes the contribution of the nonlinear interactions between the endmembers $\mathbf{m}_i$ and $\mathbf{m}_j$. The proposed bilinear models assuming (3) differ by constraints satisfied by the model parameters. Nascimento and Bioucas-Dias (2009) proposed to enforce the following constraints

$$\begin{cases} a_{r,n} \geq 0, \quad \forall r, \forall n \\ a_{r,n} \geq 0, \quad \forall r, \forall n \\ \sum_{r=1}^{R} a_{r,n} + \sum_{i=1}^{R-1}\sum_{j=i+1}^{R} \beta_{i,j,n} = 1. \end{cases} \tag{4}$$

The resulting model, referred to as Nascimento model (NM) can be seen as a linear model with additional virtual endmembers. Precisely, by considering the spectrum $\mathbf{m}_i \odot \mathbf{m}_j$ as an endmember with associated abundance $\beta_{i,j,n}$ instead of a nonlinear mixture component, the model (3) subject to (4) can be reexpressed as

$$\mathbf{y}_n = \sum_{\tilde{r}=1}^{\tilde{R}} \tilde{a}_{\tilde{r},n}\tilde{\mathbf{m}}_{\tilde{r}} + \mathbf{e}_n \tag{5}$$

where

$$\begin{cases} \tilde{a}_{\tilde{r},n} \triangleq a_{r,n}, \quad \tilde{\mathbf{m}}_{\tilde{r}} \triangleq \mathbf{m}_r, \quad \tilde{r} = 1, \ldots, R \\ \tilde{a}_{\tilde{r},n} \triangleq \beta_{i,j,n}, \quad \tilde{\mathbf{m}}_{\tilde{r}} \triangleq \mathbf{m}_i \odot \mathbf{m}_j, \quad \tilde{r} = R+1, \ldots, \tilde{R} \end{cases} \tag{6}$$

and $\tilde{R} = R(R+1)/2$. Thus the inversion step can be achieved using classical LMM-based techniques. Note that this model reduces to the LMM for $a_{\tilde{r},n} = 0, \forall \tilde{r} \in \{R+1, \ldots, \tilde{R}\}$.

A second bilinear model has been proposed by Fan et al. (2009). This model referred to as Fan model (FM) assumes $\beta_{i,j,n} = a_{i,n}a_{j,n}$ yielding

$$\mathbf{y}_n = \sum_{r=1}^{R} a_{r,n}\mathbf{m}_r + \sum_{i=1}^{R-1}\sum_{j=i+1}^{R} a_{i,n}a_{j,n}\mathbf{m}_i \odot \mathbf{m}_j + \mathbf{e}_n \tag{7}$$

subject to the constraints (2). This model is mainly motivated by the observation that if the ith component is absent in a given pixel, i.e., $a_{i,n} = 0$, there should not be nonlinear effects involving this component in the pixel, i.e., $\beta_{i,j,n} = 0, \forall j \neq i$. Moreover, if the ith component is present, the amplitudes of the nonlinear terms involving $\mathbf{m}_i$ are proportional to its abundance $a_{i,n}$. It is important to note that the FM does not generalize the LMM, contrary to the NM. More recently, Halimi et al. (2011a) proposed an FM-based bilinear model that generalizes the LMM by assuming $\beta_{i,j,n} = \gamma_{i,j,n}a_{i,n}a_{j,n}$, leading to

$$\mathbf{y}_n = \sum_{r=1}^{R} a_{r,n}\mathbf{m}_r + \sum_{i=1}^{R-1}\sum_{j=i+1}^{R} \gamma_{i,j,n}a_{i,n}a_{j,n}\mathbf{m}_i \odot \mathbf{m}_j + \mathbf{e}_n \tag{8}$$

where the nonlinearity coefficients $\gamma_{i,j,n} \in (0,1)$ allow each nonlinear effect to be quantified independently. The resulting generalized bilinear model (GBM) (8) generalizes both the LMM and the FM.

Other physics-based nonlinear mixing models

Bilinear models have been introduced to model multiple reflections between different materials. However, these models do not consider potential multiple reflections involving a single component (such as reflections occurring in tree canopy for instance). Based on a thorough physical analysis of a canyon-like situation, Meganem et al. (2013) proposed the following linear-quadratic mixing model (LQMM)

$$\mathbf{y}_n = \sum_{r=1}^{R} a_{r,n}\mathbf{m}_r + \sum_{i=1}^{R}\sum_{j=1}^{R} \beta_{i,j,n}\mathbf{m}_i \odot \mathbf{m}_j + \mathbf{e}_n \tag{9}$$

subject to the constraints (2) and $\beta_{i,j,n} \in (0,1)$. The main difference between the LQMM and the bilinear models presented above is the consideration of the quadratic terms $\mathbf{m}_r \odot \mathbf{m}_r, r = 1, \ldots, R$.

To address the modeling problem of macroscopic and microscopic mixtures, the following model has been proposed by Close et al. (2012)

$$\mathbf{y}_n = \sum_{r=1}^{R} a_{r,n}\mathbf{m}_r + a_{r+1,n}\mathcal{R}\left(\sum_{r=1}^{R} f_{r,n}\mathbf{w}_r\right) + \mathbf{e}_n \tag{10}$$

where the first term of the right-hand side of (10) is LMM-based while the second is an additional endmember weighted by the abundance $a_{r+1,n}$. This additional endmember models intimate mixtures using the average single-scattering albedo expressed in the reflectance domain through the mapping $\mathcal{R}(\cdot)$. It is worth noting that all models

presented here to model nonlinear effects do not consider dependencies between neighbor pixel, i.e, nonlinearities induced by materials in closed pixels are neglected. This assumption can seem rough but leads to simpler SU problems since the inversion step can be performed pixelwise. Recently, adjacency effects have been considered by Burazerovic et al. (2013) to address the problem of nonlinear SU assuming a bilinear model.

Model-based parametric nonlinear SU algorithms

Once a parametric model has been chosen, SU can be seen as a constrained nonlinear regression problem in the supervised case (endmembers known) or as a nonlinear blind source separation problem in the unsupervised case.

Supervised nonlinear unmixing

When the endmembers are known, most parametric model-based SU methods consist of a nonlinear regression problem that can be expressed as

$$\min_{\boldsymbol{\theta}_n \in \boldsymbol{\Theta}} \left\| \mathbf{y}_n - \psi_{\mathbf{M}} \left(\boldsymbol{\theta}_n \right) \right\|^2 \tag{11}$$

for the nth observed pixel $\mathbf{y}_n$, where $\mathbf{M} = [\mathbf{m}_1, \ldots, \mathbf{m}_R]$ is the $L \times R$ endmember matrix, $\boldsymbol{\theta}_n$ is the parameter vector of interest of the nth pixel to be estimated and $\psi_{\mathbf{M}}(\cdot)$ is a nonlinear function associated with the underlying NLMM and parameterized by $\mathbf{M}$. Moreover $\boldsymbol{\Theta}$ is the admissible set for $\boldsymbol{\theta}_n$ defined to satisfy the parameter constraints (additivity and/or positivity). The problem (11) is often tricky because it consists of a constrained nonlinear optimization problem. For the NM, it has been shown that LMM-based methods can be used to estimate the parameters of interest. For the FM and the GBM, linearization methods based on Taylor-series expansions have been proposed by Fan et al. (2009) and Halimi et al. (2011b). The GBM has received a deeper analysis since an alternative gradient-based method has also been investigated to recover the abundances and the nonlinearity parameters (Halimi et al., 2011b). To tackle convergence issues (convergence toward local optima), Halimi et al. (2011a) proposed a Bayesian algorithm based on Markov chain Monte Carlo (MCMC) methods. The Bayesian framework has the advantage of easily handling the constraints within the estimation procedure using appropriate prior distributions.

Unsupervised nonlinear unmixing

The supervised SU methods presented in the previous section assume that the endmembers are known. Geometric LMM-based EEAs have first been used to extract the endmember spectra from the nonlinearly mixed pixels. Such EEAs look for extreme points of the data clusters which often are endmembers even when nonlinear mixtures occur. However, these EEAs can suffer from the absence of pure pixels and nonlinear EEAs based on bilinear models have been studied. Gader et al. (2012) generalize the iterative SPICE algorithm (Zare and Gader, 2007) proposed for linear SU to the NM (BISPICE algorithm). Based on nonnegative matrix factorization (NMF) methods, fully unsupervised SU algorithms have also been proposed for the GBM (Yokoya et al., 2012) and the LQMM (Meganem et al., 2013). Conversely, Heylen and Scheunders (2012) derive a GBM-based EEA using geodesic distances. This recent algorithm computes distances on the manifold described by the GBM to find endmembers as extreme points of that manifold.

Model-free nonlinear SU algorithms

Nonlinear SU becomes even more challenging when the nonlinearities are unknown. Several methods have been proposed to address this problem, including approximation of geodesic distances (Heylen et al., 2011) and manifold leaning techniques (Licciardi et al., 2012; Nguyen et al., 2012). Kernel-based methods have also received growing interest over the last few years for their ability to approximate complex nonlinearities. However, if these nonparametric methods have been widely used for classification problems, their use in the unmixing context is still limited. Broadwater and Banerjee (2009); Broadwater et al. (2007) proposed methods that mainly consist of replacing the classical inner products by kernel functions. The resulting methods rely on the projection of the observed spectra onto a higher dimensional space in which the mixtures are assumed to be linear. Even if such data transformation can be very accurate for nonlinear dimensionality reduction, classification and detection, their physical interpretation is limited to address SU. More recently, nonlinear supervised algorithms based on reproducing kernel Hilbert spaces (RKHSs) have been studied to model nonlinearities resulting from interactions between endmember spectra (Chen et al., 2013b). In this scenario, the observed pixels are described by basis vectors nonlinearly related to the endmember matrix. Finally, the RKHSs have shown interesting properties for modeling various nonlinearities for supervised nonlinear unmixing, which is encouraging to move to unsupervised scenarios.

Toward nonlinearity detection in hyperspectral images

The recent contributions addressing the problem of nonlinear SU have provided interesting results in terms of mixture characterization. However, the price to pay for considering possible nonlinear effects is 1) the computational complexity that generally increases with the model complexity and 2) the degradation of the endmember and abundance estimation when assuming NLMMs in the case where the LMM is sufficient to accurately describe the mixtures. Since the LMM is often sufficient to describe most observed pixels, one possible solution consists of detecting nonlinearly from linearly mixed pixels in a pre-processing step. The pixels detected as linearly mixed could be handled subsequentially using fast and effective LMM-based methods while nonlinearly mixed pixels could be subjected to a deeper analysis. Nonlinearity detection in hyperspectral images has already been addressed by Han and Goodenough (2008) and the detection-oriented unmixing algorithm recently proposed by Dobigeon and Févotte (2013) has provided encouraging results.

On the endmember definition

As mentioned above, an endmember is assumed to be a pure spectral component to be identified in the observed image. Of course, the concept of component purity depends on the observation scale as well as the application itself. More generally, an endmember is assumed to be a component of interest. As an example, a tree can be seen either as a single endmember or as a mixture of endmembers (namely leaves and wood). Consequently, the number of endmembers and their characterization in a given scene may vary depending on the application. One of the main problems that follows from the endmember definition is the endmember variability. More precisely, two pixels composed of the same single endmember of interest can have different spectra when they do not contain the same "version" of this component. For the simple example mentioned above, the tree spectrum depends on the amount of wood and leaves contained in this component. In many applications, the endmember variability can be neglected and included in the noise modeling the model error. This assumption has been widely used in linear unmixing, which

has motivated the consideration of unique endmembers in this work. However, taking endmember variability into consideration is also an important problem, depending on the observation conditions and the observed scene (Eches et al., 2010; Somers et al., 2011; Zare et al., 2012). Even if this endmember variability has not been considered in this thesis, we think that it would deserve to be studied in future work related to nonlinear unmixing.

Structure of the manuscript

The first chapter studies a post-nonlinear mixing model (PNMM) for nonlinear SU. The proposed polynomial PNMM (PPNMM) investigated in this chapter allows nonlinearities to differ for each pixel, leading to a flexible mixture characterization. It is important to note that the nonlinearity of each pixel is characterized by a single parameter which is zero when the pixel is linearly mixed. First, supervised SU methods based on the PPNMM are considered. The first proposed algorithm is a hierarchical Bayesian algorithm coupled with MCMC methods. Two alternative optimization methods are also introduced to reduce the computational complexity of the sampling algorithm. The flexibility of the PPNMM and the performance of the three methods are then evaluated using simulations conducted on synthetic and real data. The second part of this chapter presents a new Bayesian model for unsupervised SU based on the PPNMM. In the unsupervised case, appropriate priors are assigned to the unknown endmembers to be sampled. The joint estimation of the abundances and endmembers requires a large number of parameters to be sampled. To improve the mixing properties of the sampler, constrained Hamiltonian Monte Carlo methods are investigated.

The second chapter of this manuscript considers a kernel-based approach for nonlinear SU based on a nonlinear dimensionality reduction using a *Gaussian process latent variable model* (GPLVM) (Lawrence, 2003). GPLVMs have received growing interest in the machine learning community. They have the ability to approximate various nonlinear mappings from a low-dimensional space (latent space) to a higher dimensional observation space through the use of kernel functions (Perez-Cruz et al., 2013), which makes them particularly well adapted for hyperspectral analysis and thus for SU. In this chapter, we propose to use a particular form of kernels based on the existing bilinear models, which allows the proposed unmixing strategy to be accurate when the underlying mixing model is bilinear. The first step of the proposed Bayesian algorithm performs abundance estimation. Appropriate prior distributions are assigned to the GPLVM parameters (including the latent variables related to the abundances) to be estimated and the joint posterior distribution is derived using the Bayesian inference. The maximum a posteriori (MAP) estimator of the resulting posterior is obtained using a gradient-based method and the abundance estimates are then computed. Once the abundances have been estimated, the endmembers are predicted subsequently using Gaussian process regression. The proposed unmixing algorithm breaks the usual paradigm of spectral unmixing by first estimating the abundances and then predicting the endmembers using the properties of Gaussian process (GP). Simulations conducted on synthetic and real images illustrate the flexibility of the proposed model for linear and nonlinear spectral unmixing and provide promising results for abundance and endmember estimations in spite of the absence of pure pixels in the image.

The third chapter of this manuscript is dedicated to the detection of nonlinearities in hyperspectral images. This chapter focuses on supervised detection procedures, i.e., the endmembers are assumed to be *a priori* known. The first supervised detector is based on the PPNMM studied in the first chapter. More precisely, the associated test is based on the statistical properties of the PPNMM parameters to decide whether a given pixel is linearly or

nonlinearly mixed. Conversely, the second supervised test does not assume any particular NLMM and relies on the LMM properties only. The two detectors are compared using simulations conducted on synthetic and real data. The PPNMM-based nonlinearity detector studied in the third chapter is performed after a PPNMM-based unmixing procedure and the LMM-based nonlinearity detector does not explicitly rely on an unmixing step. However, it would be interesting to derive algorithms which achieve simultaneously spectral unmixing and nonlinearity detection. The last chapter of this manuscript presents a nonlinear mixing model for joint hyperspectral image unmixing and nonlinearity detection. The proposed model assumes that the pixel reflectances are linear combinations of known pure spectral components corrupted by an additional nonlinear term, affecting the endmembers and contaminated by an additive noise. A Markov random field is considered for nonlinearity detection based on the spatial structure of the nonlinear terms. The observed image is segmented into regions where nonlinear terms, if present, share similar statistical properties. A Bayesian algorithm is proposed to estimate the parameters involved in the model yielding a joint nonlinear unmixing and nonlinearity detection algorithm. The performance of the proposed strategy is first evaluated on synthetic data. Simulations conducted with real data show the accuracy of the proposed unmixing and nonlinearity detection strategy for the analysis of hyperspectral images.

Main contributions

- **First chapter.** A new nonlinear model based on post-nonlinear mixtures is proposed for linear/nonlinear SU. The nonlinearity in each pixel is characterized by a single amplitude parameter which will allow simple nonlinearity detectors to be derived in the third chapter. SU methods are proposed to address the problems of supervised and unsupervised unmixing. Efficient sampling algorithms based on Hamiltonian MCMC methods are also used to improve the mixing properties of the sampler when the endmembers are unknown (and have to be estimated).

- **Second chapter.** A new nonparametric nonlinear SU algorithm is proposed for unsupervised SU. The proposed unmixing algorithm breaks the usual paradigm of spectral unmixing by first estimating the abundances and then predicting the endmembers in spite of the absence of pure pixel in the image.

- **Third chapter.** Statistical tests are derived for pixel-by-pixel nonlinearity detection when the endmembers are known. The proposed tests are computationally efficient and thus can be implemented in practical applications. The detection results can then be used as a pre-processing step to select pixels or regions where more complex models should be used instead of the classical LMM for SU.

- **Fourth chapter.** A new nonlinear mixing model for joint hyperspectral image unmixing and nonlinearity detection is proposed. The observed image is segmented into regions where nonlinear terms, if present, share similar statistical properties. The resulting algorithm provides accurate abundance estimates when the actual mixtures are linear and nonlinear and it thus generalizes the binary nonlinearity detectors proposed in the third chapter by considering different levels (classes) of nonlinearities.

Publications related to the PhD thesis

International journal papers

1. Y. Altmann, N. Dobigeon, S. McLaughlin and J.-Y. Tourneret, "Residual component analysis of hyperspectral images-Application to joint nonlinear unmixing and nonlinearity detection,", *IEEE Trans. Image Process.*, 2013, submitted.

2. Y. Altmann, N. Dobigeon and J.-Y. Tourneret, "Unsupervised post-nonlinear unmixing of hyperspectral images using a Hamiltonian Monte Carlo algorithm,", *IEEE Trans. Image Process.*, 2013, submitted.

3. Y. Altmann, N. Dobigeon , S. McLaughlin and J.-Y. Tourneret, "Nonlinear spectral unmixing of hyperspectral images using Gaussian processes," *IEEE Trans. on Signal Process.* 2013, vol. 61, no. 10, pp. 2442-2453, May 2013.

4. Y. Altmann, N. Dobigeon and J.-Y. Tourneret, "Nonlinearity detection in hyperspectral images using a polynomial post-nonlinear mixing model," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1267-1276, Apr. 2013.

5. Y. Altmann, A. Halimi, N. Dobigeon and J.-Y. Tourneret, "Supervised nonlinear spectral unmixing using a post-nonlinear mixing model for hyperspectral imagery," *IEEE Trans. Image Process.*, vol. 21, no. 6, pp. 3017-3025, June 2012.

International conference papers

1. Y. Altmann, N. Dobigeon , and J.-Y. Tourneret, "Bayesian algorithm for unsupervised unmixing of hyperspectral images using a post-nonlinear model," *Proc. European Signal Processing Conf. (EUSIPCO)*, to appear.

2. Y. Altmann, N. Dobigeon, J.-Y. Tourneret and S. McLaughlin, "Nonlinear hyperspectral unmixing using Gaussian processes," in *Proc. IEEE GRSS Workshop on Hyperspectral Image and SIgnal Processing: Evolution in Remote Sensing (WHISPERS)*, 2013, to appear.

3. Y. Altmann, N. Dobigeon, J.-Y. Tourneret and J. C. M. Bermudez, "A robust test for nonlinear mixture detection in hyperspectral images," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, Vancouver, Canada, 2013, to appear.

4. Y. Altmann, N. Dobigeon and J.-Y. Tourneret , "Detecting nonlinear mixtures in hyperspectral images," in *Proc. IEEE GRSS Workshop on Hyperspectral Image and SIgnal Processing: Evolution in Remote Sensing (WHISPERS)*, Shangai, China, June 2012.

5. Y. Altmann, N. Dobigeon, S. McLaughlin and J.-Y. Tourneret, "Nonlinear unmixing of hyperspectral images using Gaussian processes," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, Kyoto, Japan, March 2012, pp. 1249-1252.

6. Y. Altmann, N. Dobigeon, S. McLaughlin and J.-Y. Tourneret , "Nonlinear unmixing of hyperspectral images using radial basis functions and orthogonal least squares", in *Proc. IEEE Int.Geoscience and Remote Sensing Symp. (IGARSS)*, Vancouver, Canada, July 2011, pp. 1151-1154.

7. Y. Altmann, A. Halimi, N. Dobigeon and J.-Y. Tourneret , "A post nonlinear mixing model for hyperspectral images unmixing", in *Proc. IEEE Int.Geoscience and Remote Sensing Symp. (IGARSS)*, Vancouver, Canada, July 2011, pp. 1882-1885.

8. Y. Altmann, N. Dobigeon and J.-Y. Tourneret , "Bilinear models for nonlinear unmixing of hyperspectral images ", in *Proc. IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Lisbon, Portugal, June 2011, pp. 1-4.

9. Y. Altmann, A. Halimi, N. Dobigeon and J.-Y. Tourneret, "Supervised nonlinear spectral unmixing using a polynomial post nonlinear model for hyperspectral imagery," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011, pp. 1009-1012.

National conference papers

1. Y. Altmann, N. Dobigeon, Steve McLaughlin and J.-Y. Tourneret, "Démélange non linéaire d'images hyperspectrales à l'aide de fonctions radiales de base et de moindres carrés orthogonaux," *Actes du XXIIIième Colloque GRETSI*, Bordeaux, France, Sept. 2011, in French.

Publications related to the Master thesis

International journal papers

1. A. Halimi, Y. Altmann, N. Dobigeon and J.-Y. Tourneret, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," *IEEE Trans. Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4153-4162, Nov. 2011.

International conference papers

1. A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Unmixing hyperspectral images using the generalized bilinear model," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Vancouver, Canada, July 2011, pp. 1886-1889.

2. A. Halimi, Y. Altmann, N. Dobigeon and J.-Y. Tourneret, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," in *Proc. IEEE Workshop on Statistical Signal Processing (SSP)*, Nice, France, June 2011, pp. 413-416.

Chapter 1

Polynomial post-nonlinear mixing model for spectral unmixing

This chapter has been adapted from the journal papers Altmann et al. (2012) (published) and Altmann et al. (2013c) (submitted).

Contents

1.1 Introduction (in French)

Dans ce chapitre, nous étudions un premier modèle de mélange non-linéaire particulier, dit "post-non-linéaire polynomial" (PPNMM pour *polynomial post-nonlinear mixing model*), pour le démélange spectral. Tout d'abord, nous introduisons le modèle PPNMM et le comparons à d'autres modèles non-linéaires existants. En particulier, nous montrons que ce modèle montre de fortes similitudes avec les modèles polynomiaux de la littérature qui ont été étudiés pour modéliser les multi-trajets apparaissant en présence de relief. Un des avantages du PPNMM est sa simplicité puisqu'il met en jeu un seul paramètre supplémentaire par pixel par rapport au modèle de mélange linéaire.

Dans une deuxième étape, nous proposons trois procédures de démélange supervisé basées sur le PPNMM, c'est-à-dire lorsque les signatures spectrales des composants de l'image sont connues. La première approche se compose d'un modèle bayésien couplé avec des méthodes de simulations de Monte Carlo par chaines de Markov (MCMC). Deux méthodes d'optimisation basées sur le PPNMM sont ensuite proposées. Ces méthodes permettent d'obtenir des résultats similaires à ceux obtenus à l'aide de l'algorithme bayésien (avec un coût de calcul réduit), ce qui les rend très avantageuses pour le démélange d'images hyperspectrales de grandes tailles. Les trois méthodes de démélange supervisé sont ensuite comparées à des algorithmes de démélange linéaires et non-linéaires de la littérature par des simulations sur données synthétiques et réelles. Ces simulations permettent également d'évaluer la pertinence du PPNMM pour le démélange d'images hyperspectrales.

Dans de nombreuses applications pratiques, les signatures spectrales des composants purs présents dans l'image observée ne sont pas connues. La dernière partie de ce chapitre présente une procédure de démélange spectral bayésien totalement non-supervisée basée sur le PPNMM. Seul le nombre de composants purs de l'image est supposé connu. En raison du grand nombre de paramètres intervenant dans le problème de démélange non-supervisé, une méthode d'échantillonnage efficace basée sur des méthodes de Monte Carlo à dynamiques hamiltoniennes (HMC pour *Hamiltonian Monte Carlo*) est proposée pour estimer conjointement les signatures spectrales et les abondances de chaque pixel. Les performances de ce dernier algorithme de démélange sont évaluées sur données synthétiques et réelles, de manière similaire à l'étude menée dans le cas supervisé. En particulier, ces simulations montrent les avantages de l'estimation conjointe des signatures spectrales et des abondances de chaque pixel.

1.2 Introduction

In this chapter, we study a particular nonlinear mixing model for SU referred to as *polynomial post-nonlinear mixing model* (PPNMM). First, we introduce the PPNMM and compare it to other existing nonlinear models. In a second step, we derive three PPNMM-based unmixing procedures for supervised unmixing, i.e., when the endmembers are known. The first approach consists of a Bayesian model coupled with Markov Chain Monte Carlo (MCMC) methods. Two PPNMM-based optimization methods are then proposed to reduce the computation complexity of the sampling procedure associated with the Bayesian algorithm. These optimization methods provide results similar to the Bayesian algorithm with a reduced computational cost, making them very attractive for hyperspectral image unmixing. In many practical applications, the endmembers present in the observed image are unknown. The last part of this chapter presents a fully unsupervised Bayesian unmixing procedure based on the PPNMM. Due to the large number of parameters involved in the unsupervised SU problem, an efficient sampling procedure based on Hamiltonian dynamics is proposed, leading to a Hamiltonian Monte Carlo (HMC) simulation method. Results obtained on synthetic and real images illustrate the flexibility and accuracy of the PPNMM for supervised and unsupervised SU and the performance of the corresponding estimation algorithms.

1.3 Polynomial Post-Nonlinear Mixing Model

This section defines the PPNMM used for hyperspectral image SU. Consider a set of N pixels observed at L different spectral bands. The spectrum of the nth mixed pixel $\mathbf{y}_n = [y_{1,n}, \ldots, y_{L,n}]^T$ is defined as a nonlinear transformation $\mathbf{g}_n$ of a linear mixture of R spectra $\mathbf{m}_r$ contaminated by additive noise

$$\mathbf{y}_n = \mathbf{g}_n \left(\sum_{r=1}^{R} a_{r,n} \mathbf{m}_r \right) + \mathbf{e}_n = \mathbf{g}_n \left(\mathbf{M} \boldsymbol{a}_n \right) + \mathbf{e}_n \tag{1.1}$$

where $\mathbf{m}_r = [m_{r,1}, \ldots, m_{r,L}]^T$ is the spectrum of the rth material present in the scene, $a_{r,n}$ is its corresponding proportion in the nth pixel of the image, R is the number of endmembers contained in the image, $\mathbf{g}_n(\cdot)$ is an appropriate nonlinear function depending on the nth pixel and $\mathbf{e}_n$ is an additive noise sequence. Note that the usual matrix and vector notations $\mathbf{M} = [\mathbf{m}_1, \ldots, \mathbf{m}_R]$ and $\boldsymbol{a}_n = [a_{1,n}, \ldots, a_{R,n}]^T$ have been used in the right hand side of (1.1).

The choice of the nonlinearity $\mathbf{g}_n$ deserves a specific attention. Polynomials, sigmoidal functions and combinations of polynomial and sigmoidal nonlinearities have shown interesting properties for source separation (Babaie-Zadeh et al., 2001; Jutten and Karhunen, 2003). This chapter focuses on second order polynomial nonlinearities defined by

$$\begin{array}{rccc} \mathbf{g}_n : & [0,1]^L & \to \mathbb{R}^L & \\ & \mathbf{s} & \mapsto & \begin{bmatrix} s_1 + b_n s_1^2 \\ \vdots \\ s_L + b_n s_L^2 \end{bmatrix} \end{array} \tag{1.2}$$

with $\mathbf{s} = [s_1, \ldots, s_L]^T$. An interesting property of the resulting nonlinear model referred to as PPNMM is that it reduces to the classical LMM for $b_n = 0$. Thus, we can expect unmixing results at least as good as those presented

by Dobigeon et al. (2008) and Heinz and C.-I Chang (2001) where Bayesian and least-squares (LS) methods were investigated for SU assuming the LMM. Another motivation for using the PPNMM is the Weierstrass approximation theorem which states that any continuous function defined on a bounded interval can be uniformly approximated by a polynomial with any desired precision (Mathews and Sicuranza, 2000, p. 15). As explained by Nascimento and Bioucas-Dias (2009), it is reasonable to consider polynomials with first and second order terms (since higher order terms can generally be neglected) which leads to (1.2). Higher order terms could be considered in the presence of more than two reflections (bilinear models also consider only at most two reflections). However, the resulting interaction spectra are in practice of low amplitude and are hardly distinguishable from the noise. Straightforward computations allow the PPNMM observation vector (for a given pixel of the image) to be expressed as follows

$$\mathbf{y}_n = \mathbf{g}_n\left(\mathbf{M}\boldsymbol{a}_n\right) + \mathbf{e}_n = \mathbf{M}\boldsymbol{a}_n + b_n(\mathbf{M}\boldsymbol{a}_n) \odot (\mathbf{M}\boldsymbol{a}_n) + \mathbf{e}_n \tag{1.3}$$

where $\odot$ denotes the Hadamard (term-by-term) product. Note that the resulting PPNMM includes bilinear terms such as those considered by Fan et al. (2009); Halimi et al. (2011a); Nascimento and Bioucas-Dias (2009); Somers et al. (2009). However, the nonlinear terms are characterized by a single amplitude parameter b_n, leading to a less complex model when compared with the models introduced by Nascimento and Bioucas-Dias (2009); Somers et al. (2009) and Halimi et al. (2011a). Note that the endmember $\mathbf{m}_r$ (contained in the matrix $\mathbf{M}$) can be obtained from (1.3) in the noise free case ($\mathbf{e}_n = \mathbf{0}_L$) by setting $b_n = 0$ and $\boldsymbol{a}_n = [\mathbf{0}_{r-1}, 1, \mathbf{0}_{R-r}]^T$ in (1.3).
Due to physical considerations, each abundance vector $\boldsymbol{a}_n, n \in \{1, \ldots, N\}$ satisfies the following positivity and sum-to-one constraints

$$\begin{cases} a_{r,n} \geq 0, \quad \forall r \in \{1, \ldots, R\} \\ \sum_{r=1}^{R} a_{r,n} = 1. \end{cases} \tag{1.4}$$

In this study we consider the sum-to-one constraints for the abundances. However they could be omitted as in (Meganem et al., 2013). Moreover, each endmember $\mathbf{m}_r = [m_{r,1}, \ldots, m_{r,L}]^T, r \in \{1, \ldots, R\}$ is a reflectance vector satisfying the following constraints

$$0 \leq m_{r,\ell} \leq 1, \quad \forall \ell \in \{1, \ldots, L\}. \tag{1.5}$$

1.4 Supervised PPNMM-based unmixing

In this first scenario, we assume that the endmembers contained in the hyperspectral image are known. The three PPNMM-based supervised SU algorithms presented in this manuscript allow the image pixels to be unmixed independently. Thus, for ease of reading, all subscripts referring to pixel indexes will be omitted in this section. Let $\mathbf{y}$ be the $L \times 1$ observed pixel satisfying

$$\mathbf{y} = \mathbf{M}\boldsymbol{a} + b(\mathbf{M}\boldsymbol{a}) \odot (\mathbf{M}\boldsymbol{a}) + \mathbf{e}. \tag{1.6}$$

The noise sequence $\mathbf{e}$ is an additive independent and identically distributed (i.i.d) zero-mean Gaussian noise sequence with variance σ^2, denoted as $\mathbf{e} \sim \mathcal{N}\left(\mathbf{0}_L, \sigma^2 \mathbf{I}_L\right)$, where $\mathbf{I}_L$ is the $L \times L$ identity matrix. Since the endmember matrix $\mathbf{M}$ is assumed to be known, the only parameters to be estimated are the abundance vector $\boldsymbol{a}$, the nonlinearity parameter b and the noise variance σ^2.

The supervised unmixing problem is identifiable since the application

$$\mathbf{g}: \quad \mathbb{R}^R \times \mathbb{R} \quad \to \mathbb{R}^L \tag{1.7}$$

$$(\boldsymbol{a}, b) \quad \mapsto \mathbf{M}\boldsymbol{a} + b(\mathbf{M}\boldsymbol{a}) \odot (\mathbf{M}\boldsymbol{a}) \tag{1.8}$$

is injective under specific conditions related to the pure component spectra. Note however that the function $\mathbf{s} \mapsto \mathbf{g}(\mathbf{s})$ in (1.2) (where the variable b is fixed) is non-injective (see Appendix A for details).

1.4.1 Bayesian estimation

The first proposed algorithm for supervised SU using the PPNMM consists of a Bayesian model coupled with MCMC methods. This model generalizes the hierarchical Bayesian model introduced by Dobigeon et al. (2008) for linear SU to the PPNMM. The unknown parameter vector associated with the PPNMM contains the pixel abundances $\boldsymbol{a}$, the nonlinearity parameter b and the additive noise variance σ^2. In the Bayesian framework, appropriate prior distributions are chosen for the unknown parameters. The joint posterior distribution of these parameters is then derived. However, the classical Bayesian estimators cannot be easily computed from this joint posterior. To alleviate this problem, an MCMC method is used to generate samples according to the posterior of interest. The generated samples are then used to approximate the Bayesian estimators. As in any Bayesian algorithm, the joint posterior distribution can also be used to compute confidence intervals for the parameter estimates.

Likelihood

Equation (1.6) shows that $\mathbf{y}|\boldsymbol{a}, b, \sigma^2$ is distributed according to a Gaussian distribution with mean $\mathbf{g}(\mathbf{M}\boldsymbol{a})$ and covariance matrix $\sigma^2\mathbf{I}_L$, denoted as $\mathbf{y}|\boldsymbol{a}, b, \sigma^2 \sim \mathcal{N}\left(\mathbf{g}(\mathbf{M}\boldsymbol{a}), \sigma^2\mathbf{I}_L\right)$. As a consequence, the likelihood function of the observation vector $\mathbf{y}$ can be expressed as

$$f(\mathbf{y}|\boldsymbol{a}, b, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{L}{2}} \exp\left(-\frac{\|\mathbf{y} - \mathbf{g}(\mathbf{M}\boldsymbol{a})\|^2}{2\sigma^2}\right) \tag{1.9}$$

where $\|\mathbf{x}\| = \sqrt{\mathbf{x}^T\mathbf{x}}$ is the standard ℓ^2 norm.

Parameter priors

In order to satisfy the sum-to-one constraint, the abundance vector can be rewritten[1] $\boldsymbol{a} = [\mathbf{c}, a_R]^T$ where $\mathbf{c} = [a_1, \ldots, a_{R-1}]^T$, $a_R = 1 - \sum_{r=1}^{R-1} a_r$. The positivity constraints in (1.4) impose that $\mathbf{c}$ belongs to the following simplex $\mathcal{S}$

$$\mathcal{S} = \left\{ \mathbf{c} = [c_1, \ldots, c_{R-1}]^T \middle| c_r \geq 0, \forall r < R, \sum_{r=1}^{R-1} c_r \leq 1 \right\}. \tag{1.10}$$

A uniform prior distribution on $\mathcal{S}$ is chosen for $\mathbf{c}$ to reflect the absence of prior knowledge about the abundance vector. Note that choosing this prior for $\mathbf{c}$ is equivalent to choosing a Dirichlet prior distribution with all hyperparameters set to 1.

A Jeffreys' prior is chosen for σ^2

$$f(\sigma^2) \propto \frac{1}{\sigma^2}\mathbf{I}_{\mathbb{R}^+}(\sigma^2) \tag{1.11}$$

[1] Note that the proposed parametrization is chosen for notation simplicity. However, the component to be discarded can be randomly chosen.

which also reflects the absence of knowledge for this parameter (see (Punskaya et al., 2002) for details).
A conjugate Gaussian prior is finally chosen for the nonlinearity parameter b

$$b\,|\sigma_b^2 \sim \mathcal{N}\left(0, \sigma_b^2\right). \tag{1.12}$$

The Gaussian prior is zero-mean since the value of b can be equally likely positive or negative. Moreover, it favors small values of b and is a conjugate prior for the parameter b which will simplify the computations.

Hyperparameter prior

The performance of the proposed Bayesian model for spectral unmixing depends on the values of the hyperparameter σ_b^2. When hyperparameters are difficult to adjust, it is classical to include them in the unknown parameter vector, resulting in a hierarchical Bayesian model (Robert, 2007). This strategy requires to define a prior distribution for the hyperparameter σ_b^2. A conjugate inverse-gamma prior is assigned to σ_b^2

$$\sigma_b^2 \sim \mathcal{IG}\left(\gamma, \nu\right) \tag{1.13}$$

where (γ, ν) are real parameters fixed to obtain a flat prior, reflecting the absence of knowledge about the variance σ_b^2 ((γ, ν) will be set to $(1, 10^{-2})$ in the simulation section). The resulting directed acyclic graph (DAG) is depicted in Fig. 1.1.

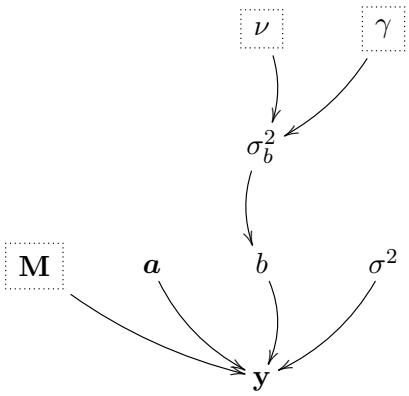


Figure 1.1: DAG for the parameter priors and hyperpriors (the fixed parameters appear in dashed boxes).

Joint posterior distribution

The joint posterior distribution of the unknown parameter/hyperparameter vector $\left\{\boldsymbol{\theta}, \sigma_b^2\right\}$, where $\boldsymbol{\theta} = \left\{\mathbf{c}, b, \sigma^2\right\}$ can be computed using the following hierarchical structure

$$f(\boldsymbol{\theta}, \sigma_b^2|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\sigma_b^2) f(\sigma_b^2) \tag{1.14}$$

where $\propto$ means "proportional to" and $f(\mathbf{y}|\boldsymbol{\theta})$ is defined in (1.9). By assuming the parameters σ^2, b and $\boldsymbol{a}_{\setminus R}$ are *a priori* independent, the joint prior distribution of the unknown parameter vector can be expressed as

$$f(\boldsymbol{\theta}|\sigma_b^2) \quad = \quad f(\mathbf{c}) f(\sigma^2) f(b|\sigma_b^2). \tag{1.15}$$

The joint posterior distribution $f(\boldsymbol{\theta}, \sigma_b^2|\mathbf{y})$ can then be computed up to a multiplicative constant

$$f(\boldsymbol{\theta}, \sigma_b^2|\mathbf{y}) \propto \frac{1}{\sigma^2}\left(\frac{1}{\sigma_b^2}\right)^{\frac{3}{2}+\gamma} \exp\left(-\frac{b^2+2\nu}{2\sigma_b^2}\right) f(\mathbf{y}|\mathbf{c}, \sigma^2, b)\mathbf{1}_{\mathcal{S}}(\mathbf{c}). \tag{1.16}$$

Unfortunately, it is difficult to obtain closed form expressions of the standard Bayesian estimators (including the maximum a posteriori (MAP) and the minimum mean square error (MMSE) estimators) associated with (1.16). The last part of this section studies a Markov chain Monte Carlo (MCMC) method which can be used to generate samples asymptotically distributed according to (1.16). These generated samples are then used to compute the MAP or MMSE estimators of the unknown parameters $\{\boldsymbol{\theta}, \sigma_b^2\}$.

Metropolis-within-Gibbs sampler

The principle of the Gibbs sampler is to sample according to the conditional distributions of the posterior of interest (Robert and Casella, 2004, Chap. 10). The conditional distributions associated with the posterior (1.16) are studied below.

Sampling from $f(c_r|\mathbf{y}, \boldsymbol{\theta}_{\backslash c_r}, \sigma_b^2)$ Straightforward computations lead to

$$f(c_r|\mathbf{y}, \boldsymbol{\theta}_{\backslash c_r}, \sigma_b^2) \propto \exp\left(-\frac{\|\mathbf{y} - \mathbf{g}\left(\mathbf{M}\boldsymbol{a}\right)\|^2}{2\sigma^2}\right) \mathbf{1}_{\mathcal{S}}(\mathbf{c}) \tag{1.17}$$

where $r = 1, \ldots, R-1$ and where the notation $\boldsymbol{\theta}_{\backslash c_r}$ indicates that c_r has been removed from the vector $\boldsymbol{\theta}$. Mainly because of the indicator function $\mathbf{1}_{\mathcal{S}}(\mathbf{c})$ and the nonlinear function $\mathbf{g}(\cdot)$, it is not easy to sample according to (1.17). Thus, we propose to update the coefficient c_r thanks to a Metropolis-Hasting move. More precisely, a new abundance coefficient is proposed following a truncated Gaussian random walk procedure to satisfy the constraints (1.4) (the variance of the proposal distribution has been adjusted to obtain an acceptance rate close to 0.5, as recommended in (Robert and Cellier, 1998, p. 8)). The generated sampler is accepted or rejected with an appropriate probability ρ provided in Algo. 1.1.

Sampling from $f(b|\mathbf{y}, \boldsymbol{\theta}_{\backslash b}, \sigma_b^2)$ Using (1.16), it can be easily shown that b is distributed according to the following Gaussian distribution

$$b|\mathbf{y}, \boldsymbol{\theta}_{\backslash b}, \sigma_b^2 \sim \mathcal{N}\left(m_b, s_b^2\right) \tag{1.18}$$

where

$$m_b = \frac{\sigma_b^2\left(\mathbf{y} - \mathbf{M}\boldsymbol{a}\right)^T \mathbf{h}(\boldsymbol{a})}{\sigma_b^2\mathbf{h}(\boldsymbol{a})^T\mathbf{h}(\boldsymbol{a}) + \sigma^2}, \quad s_b^2 = \frac{\sigma_b^2\sigma^2}{\sigma_b^2\mathbf{h}(\boldsymbol{a})^T\mathbf{h}(\boldsymbol{a}) + \sigma^2}$$

and $\mathbf{h}(\boldsymbol{a}) = (\mathbf{M}\boldsymbol{a}) \odot (\mathbf{M}\boldsymbol{a})$. As a consequence, sampling according to (1.18) is straightforward.

Sampling from $f(\sigma^2|\mathbf{y}, \boldsymbol{\theta}_{\backslash \sigma^2}, \sigma_b^2)$ By considering the posterior distribution (1.16), it can be shown that $\sigma^2|\mathbf{y}, \boldsymbol{\theta}_{\backslash \sigma^2}, \sigma_b^2$ is distributed according to the following inverse-gamma distribution

$$\sigma^2|\mathbf{y}, \boldsymbol{\theta}_{\backslash \sigma^2}, \sigma_b^2 \sim \mathcal{IG}\left(\frac{L}{2}, \frac{\|\mathbf{y} - \mathbf{g}(\mathbf{M}\boldsymbol{a})\|^2}{2}\right) \tag{1.19}$$

from which it is easy to sample.

1: Initialization ($t = 0$)

- $\boldsymbol{a}^{(0)}, b^{(0)}, \sigma^{2(0)}, \sigma_b^{2(0)}$.

2: Iterations ($t \geq 1$)

3: Set $\tilde{\mathbf{c}} = [\tilde{c}_1, \ldots, \tilde{c}_{R-1}]^T = \mathbf{c}^{(t-1)}$

4: **for** $r = 1 : R-1$ **do**

5:

- Sample a candidate ζ_r using a Gaussian proposal distribution $\mathcal{N}_{\left[0,1-\tilde{\mathbf{c}}_{\backslash r}^T \mathbf{1}_{R-2}\right]}\left(\tilde{c}_r, \sigma_r^2\right)$.
- Compute $\rho = \min\left\{ \dfrac{f(\zeta_r|\mathbf{y}, b, \tilde{\mathbf{c}}_{\backslash r}, \sigma^2, \sigma_b^2)}{f(\tilde{c}_r|\mathbf{y}, b, \tilde{\mathbf{c}}_{\backslash r}, \sigma^2, \sigma_b^2)} \dfrac{\left[\Phi\left((1 - \tilde{\mathbf{c}}_{\backslash r}^T \mathbf{1}_{R-2} - \zeta_r)/\sigma_r\right) - \Phi\left(-\zeta_r/\sigma_r\right)\right]}{\left[\Phi\left((1 - \tilde{\mathbf{c}}_{\backslash r}^T \mathbf{1}_{R-2} - \tilde{c}_r)/\sigma_r\right) - \Phi\left(-\tilde{c}_r/\sigma_r\right)\right]}, 1 \right\}$
- Set $\tilde{c}_r = \begin{cases} \zeta_r & \text{with probability} \quad \rho \\ \tilde{c}_r & \text{with probability} \quad 1-\rho \end{cases}$

6: **end for**

7: Set $\mathbf{c}^{(t)} = [\tilde{c}_1, \ldots, \tilde{c}_{R-1}]^T$

8: Set $a_R^{(t)} = 1 - \sum_{r=1}^{R-1} a_r^{(t)}$

9: Sample $b^{(t)}$ from the pdf in (1.18)

10: Sample $\sigma^{2(t)}$ from the pdf in (1.19)

11: Sample $\sigma_b^{2(t)}$ from the pdf in (1.20)

12: Set $t = t + 1$.

ALGO. 1.1: Gibbs Sampling Algorithm

Sampling from $f(\sigma_b^2|\mathbf{y}, \boldsymbol{\theta})$ Finally, by looking at the posterior distribution (1.16), it can be seen that $\sigma_b^2|\mathbf{y}, \boldsymbol{\theta}$ is distributed according to the following inverse-gamma distribution

$$\sigma_b^2|\mathbf{y}, \boldsymbol{\theta} \sim \mathcal{IG}\left(\frac{1}{2} + \gamma, \frac{b^2}{2} + \nu\right). \tag{1.20}$$

The resulting Metropolis-within-Gibbs sampler used to sample according to the posterior (1.16) is summarized in Algo. 1.1.

After generating samples using the procedures detailed above, the MMSE estimator of the unknown parameters can be approximated by computing the empirical averages of these samples, after an appropriate burn-in period[2]. Even if the sampling strategy has been observed to converge very fast, its computational complexity can be heavy for some practical applications. The next section studies least squares estimators which allow this computational complexity to be significantly reduced.

[2]The length of the burn-in period has been determined using appropriate convergence diagnoses (Robert and Cellier, 1998).

1.4.2 Least squares methods

Least squares (LS) methods have been used successfully for linear unmixing (Heinz and C.-I Chang, 2001). The LS methods associated with the observation equation (1.6) consist of minimizing the following criterion

$$\begin{aligned} J(\boldsymbol{a}, b) &= \frac{1}{2} \|\mathbf{y} - \mathbf{g}(\mathbf{M}\boldsymbol{a})\|^2 \\ &= \frac{1}{2} \|\mathbf{y} - \mathbf{M}\boldsymbol{a} - b(\mathbf{M}\boldsymbol{a}) \odot (\mathbf{M}\boldsymbol{a})\|^2 \end{aligned} \tag{1.21}$$

under the positivity and sum-to-one constraints (1.4) for the abundance vector. This optimization problem is not easy to handle mainly because of the constraints (1.4). However, the cost function $J(\boldsymbol{a}, b)$ is quadratic with respect to the parameter b. As a consequence, by differentiating $J(\boldsymbol{a}, b)$ with respect to b, the following closed-form expression for b can be obtained

$$b = \frac{(\mathbf{y} - \mathbf{M}\boldsymbol{a})^T \mathbf{h}(\boldsymbol{a})}{\mathbf{h}(\boldsymbol{a})^T \mathbf{h}(\boldsymbol{a})} = \check{b}(\boldsymbol{a}). \tag{1.22}$$

After replacing (1.22) in $J(\boldsymbol{a}, b)$, we obtain[3]

$$J(\boldsymbol{a}) = J\left[\boldsymbol{a}, \check{b}\left(\boldsymbol{a}\right)\right] = \frac{1}{2} \|\mathbf{y} - \check{\mathbf{y}}(\boldsymbol{a})\|^2 \tag{1.23}$$

where

$$\check{\mathbf{y}}(\boldsymbol{a}) = \mathbf{M}\boldsymbol{a} + \check{b}(\boldsymbol{a})(\mathbf{M}\boldsymbol{a}) \odot (\mathbf{M}\boldsymbol{a}). \tag{1.24}$$

We introduce below two strategies to compute the optimal abundance vector

$$\hat{\boldsymbol{a}} = \arg\min_{\boldsymbol{a}} J(\boldsymbol{a})$$

under the constraints (1.4). Note that once $\hat{\boldsymbol{a}}$ has been computed, the nonlinearity parameter b can be estimated as follows

$$\hat{b} = \check{b}\left(\hat{\boldsymbol{a}}\right), \tag{1.25}$$

and the noise variance can be subsequently estimated using

$$\hat{\sigma}^2 = \frac{1}{L} \left\|\mathbf{y} - \mathbf{M}\hat{\boldsymbol{a}} - \hat{b}(\mathbf{M}\hat{\boldsymbol{a}}) \odot (\mathbf{M}\hat{\boldsymbol{a}})\right\|^2. \tag{1.26}$$

Taylor approximation

Motivated by the method introduced by Fan et al. (2009), we propose to approximate the function $\check{\mathbf{y}}(\cdot)$ defined in (1.24) using a Taylor series expansion where only first-order terms are considered. Let $\boldsymbol{a}^{(t)}$ denotes the estimated abundance vector estimate at the tth step of the proposed iterative algorithm, and its corresponding estimated spectrum $\check{\mathbf{y}}(\boldsymbol{a}^{(t)})$ following (1.24). The Taylor approximation of $\check{\mathbf{y}}(\cdot)$ at $\boldsymbol{a}^{(t)}$ can be written

$$\check{\mathbf{y}}(\boldsymbol{a}) \approx \check{\mathbf{y}}\left(\boldsymbol{a}^{(t)}\right) + \nabla\check{\mathbf{y}}\left(\boldsymbol{a}^{(t)}\right)\left(\boldsymbol{a} - \boldsymbol{a}^{(t)}\right) \tag{1.27}$$

where $\nabla\check{\mathbf{y}}(\boldsymbol{a}^{(t)})$ is the gradient matrix of $\check{\mathbf{y}}(\boldsymbol{a}^{(t)})$ of size $L \times R$ and $\boldsymbol{a}$ is the unknown parameter vector to be estimated. The rth column of the gradient matrix $\nabla\check{\mathbf{y}}(\boldsymbol{a}^{(t)})$ can be derived from (1.24)

$$\frac{\partial\check{\mathbf{y}}(\boldsymbol{a})}{\partial a_r} = \mathbf{m}_r + \frac{\partial\check{b}(\boldsymbol{a})}{\partial a_r}\mathbf{h}(\boldsymbol{a}) + \check{b}(\boldsymbol{a})\frac{\partial\mathbf{h}(\boldsymbol{a})}{\partial a_r} \tag{1.28}$$

[3] For brevity, the same notation J is chosen for the criteria depending on $\boldsymbol{a}$ and $(\boldsymbol{a}, b)$.

where $r = 1, \ldots, R$ and the partial derivatives of $\check{b}(\cdot)$ and $\mathbf{h}(\cdot)$ are available in Appendix B. Approximating $\check{\mathbf{y}}(\cdot)$ in (1.23) using (1.27), the vector $\boldsymbol{a}^{(t+1)}$ can then be estimated by solving the following constrained LS problem

$$\boldsymbol{a}^{(t+1)} = \arg\min_{\boldsymbol{a}} \left\| \mathbf{s}^{(t)} - \tilde{\mathbf{M}}^{(t)} \boldsymbol{a} \right\|^2, \tag{1.29}$$

under the constraints (1.4), where

$$\mathbf{s}^{(t)} = \mathbf{y} - \check{\mathbf{y}}\left(\boldsymbol{a}^{(t)}\right) + \nabla\check{\mathbf{y}}\left(\boldsymbol{a}^{(t)}\right) \boldsymbol{a}^{(t)} \tag{1.30}$$

and $\tilde{\mathbf{M}}^{(t)} = \nabla\check{\mathbf{y}}\left(\boldsymbol{a}^{(t)}\right)$ is the $L \times R$ gradient matrix. Problem (1.29) can finally be solved by the FCLS algorithm (Heinz and C.-I Chang, 2001). More precisely, the sum-to-one constraint of the abundances is considered by penalizing (1.29), leading to

$$\boldsymbol{a}^{(t+1)} = \arg\min_{\boldsymbol{a}} \left[\left\| \mathbf{s}^{(t)} - \tilde{\mathbf{M}}^{(t)} \boldsymbol{a} \right\|^2 + \eta(1 - \mathbf{1}_R^T \boldsymbol{a})^2 \right] \tag{1.31}$$

subject to the non-negativity constraints for the parameter vector $\boldsymbol{a}$, where $\eta \in \mathbb{R}^+$ controls the impact of the sum-to-one constraint. The procedure (1.31) is repeated until convergence and is summarized in Algo. 1.2. The convergence of this iterative procedure to the global minimum of the objective function (1.25) is difficult to prove because of the constraints (1.4) in (1.29). The next section introduces an alternative subgradient-based algorithm whose convergence (to a local minimum of the associated objective function) is ensured.

1: Initialization ($t = 0$)

- Set $\boldsymbol{a}^{(0)}$

2: Iterations ($t \geq 0$)
3: Compute the gradient matrix of $\check{\mathbf{y}}$ at $\boldsymbol{a}^{(t)}$ using (1.28)
4: Compute $\boldsymbol{a}^{(t)}$ using (1.29)
5: Compute $b^{(t)}$ using (1.25)
6: Set $t = t + 1$.

ALGO. 1.2: Taylor Approximation Algorithm

Subgradient-based optimization

A classical gradient approach could be used to solve the cost function defined in (1.23) in absence of constraints. However, the problem is more complicated when the constraints (1.4) have to be considered. The estimation method studied in this section is based on a subgradient optimization (SO) algorithm (Bazaraa et al., 1993, p. 339) that is appropriate for constrained problems. More precisely, subgradient-based optimization allows each abundance $a_1, \ldots, a_r$ to be updated independently. Thanks to the sum-to-one constraint of the abundance vector, the cost function (1.23) can be expressed as a function of $\mathbf{c} = [a_1, \ldots, a_{R-1}]^T$ by setting $a_R = 1 - \sum_{r=1}^{R-1} a_r$. In that case, the cost function (1.23) can be rewritten

$$J^*(\mathbf{c}) = \frac{1}{2} \left\| \mathbf{y} - \check{\mathbf{y}}^*(\boldsymbol{a}_{\setminus R}) \right\|^2 \tag{1.32}$$

where

$$J^*(\mathbf{c}) = J\left(a_1, \ldots, a_{R-1}, 1 - \sum_{r=1}^{R-1} a_r\right) \tag{1.33}$$

$$\check{\mathbf{y}}^*(\mathbf{c}) = \check{\mathbf{y}}\left(a_1, \ldots, a_{R-1}, 1 - \sum_{r=1}^{R-1} a_r\right). \tag{1.34}$$

1: Initialization ($t = 0$)

- Set $\boldsymbol{a}^{(0)}$

2: Iterations ($t \geq 1$)
3: Set $\tilde{\mathbf{c}} = [\tilde{c}_1, \ldots, \tilde{c}_{r-1}]^T = \mathbf{c}^{(t-1)} = [a_1^{(t-1)}, \ldots, a_{R-1}^{(t-1)}]^T$
4: **for** $r = 1 : R-1$ **do**
5:
- Compute $d_r = -\frac{\partial J^*(\tilde{\mathbf{c}})}{\partial \tilde{c}_r}$
- Compute $\lambda_{r,M}$ from (1.36)
- Compute $\hat{\lambda}_r$ from (1.35)
- Set $\tilde{c}_r = \tilde{c}_r - \hat{\lambda}_r d_r$

6: **end for**
7: Set $\mathbf{c}^{(t)} = \tilde{\mathbf{c}}$
8: Set $a_R^{(t)} = 1 - \sum_{r=1}^{R-1} c_r^{(t)}$
9: Compute $b^{(t)}$ using (1.25)
10: Set $t = t + 1$.

ALGO. 1.3: Constrained Subgradient Algorithm

At a given point $\mathbf{c}$, the SO algorithm performs sequential line searches along the directions d_r defined by the partial derivatives with respect to c_r (for $r = 1, \ldots, R-1$), i.e.,

$$d_r = -\frac{\partial J^*(\mathbf{c})}{\partial c_r} = [\mathbf{y} - \check{\mathbf{y}}^*(\mathbf{c})]^T \frac{\partial \check{\mathbf{y}}^*(\mathbf{c})}{\partial c_r}$$

where the partial derivatives of $\check{\mathbf{y}}^*(\boldsymbol{a}_{\setminus R})$ are provided in Appendix B. Finally, the line search procedure solves the following problem

$$\hat{\lambda}_r \quad = \quad \arg\min_{\lambda_r} \quad J^*(\mathbf{c} - \lambda_r \mathbf{u}_r) \tag{1.35}$$

where $\mathbf{u}_r = [0, \ldots, \text{sign}(d_r), 0, \ldots, 0]^T$ is a direction vector of size $(R-1) \times 1$, $0 \leq \lambda_r \leq \lambda_{r,M}$ and $\lambda_{r,M} \in \mathbb{R}^+$ (for $r = 1, \ldots, R-1$) are upper bounds for the line search parameters. More precisely, upper bounding λ_r according to the following rule

$$\lambda_{r,M} = \begin{cases} 0, & \text{if } \; d_r = 0 \\ c_r, & \text{if } \; d_r > 0 \\ c_r - \sum_{i=1, i \neq r}^{R-1} c_i, & \text{if } \; d_r < 0 \end{cases}$$

ensures the constraints (1.4) are satisfied. The problem (1.35) can be solved using the golden section method (Bazaraa et al., 1993, p. 270). The abundances are then updated component by component. Here again, the procedure is repeated until convergence. The final algorithm is summarized in Algo. 1.3. The next section presents the performance of the proposed algorithms on synthetic and real hyperspectral images.

1.4.3 Simulations

Synthetic data

The performance of the proposed nonlinear SU algorithms is first evaluated by unmixing 4 synthetic images of size 50×50 pixels. The $R = 3$ endmembers contained in these images have been extracted from the spectral libraries provided with the ENVI software (RSI (Research Systems Inc.), 2003) (i.e., green grass, olive green paint and galvanized steel metal). The first synthetic image I_1 has been generated using the standard linear mixing model (LMM). A second image I_2 has been generated according to the bilinear mixing model introduced by Fan et al. (2009), referred to as "Fan model" (FM). A third image I_3 has been generated according to the generalized bilinear mixing model (GBM) presented by Halimi et al. (2011a), whereas a fourth image I_4 has been generated according to the PPNMM. For each image, the abundance vectors $\boldsymbol{a}_n, p = 1, \ldots, 2500$ have been randomly generated according to a uniform distribution over the admissible set defined by the positivity and sum-to-one constraints. All images have been corrupted by an additive white Gaussian noise of variance $\sigma^2 = 2.8 \times 10^{-3}$, corresponding to a signal-to-noise ratio SNR $= L^{-1}\sigma^{-2} \|\mathbf{g}(\boldsymbol{a})\|^2 \simeq 15$dB. The nonlinearity coefficients are uniformly drawn in the set $(0, 1)$ for the GBM and the parameter b has been generated uniformly in the set $(-0.3, 0.3)$ for the PPNMM. Different estimation procedures have been considered for the four mixing models. More precisely,

- for the LMM, we have considered the standard FCLS algorithm (Heinz and C.-I Chang, 2001) and the Bayesian algorithm by Dobigeon et al. (2008),
- the FM has been unmixed using the LS method introduced by Fan et al. (2009) and a Bayesian algorithm similar to the one derived by Halimi et al. (2011a) but assuming all the nonlinearity coefficients are equal to 1,
- the unmixing strategies used for the GBM are the three algorithms presented in (Halimi et al., 2011b), i.e., a Bayesian algorithm and two LS methods,
- the Bayesian and LS algorithms presented in this chapter have been used for unmixing the proposed PPNMM. Note that all results presented in this study have been obtained using the Bayesian MMSE estimator.

The quality of the unmixing procedures can be measured by comparing the estimated and actual abundance vector using the root normalized mean square error (RNMSE) defined by

$$\text{RNMSE} = \sqrt{\frac{1}{NR}\sum_{n=1}^{N} \|\hat{\boldsymbol{a}}_n - \boldsymbol{a}_n\|^2} \tag{1.36}$$

where $\boldsymbol{a}_n$ is the nth actual abundance vector and $\hat{\boldsymbol{a}}_n$ its estimate. Table 1.1 shows the RNMSEs associated with the images $I_1, \ldots, I_4$ for the different estimation procedures. Note that the best results (in term of RNMSE) for each image have been represented in bold and blue whereas the second best results have been depicted in bold.

Table 1.1: Abundance RNMSEs ($\times 10^{-2}$): synthetic images .

		I_1 (LMM)	I_2 (FM)	I_3 (GBM)	I_4 (PPNMM)
LMM	Bayesian (Dobigeon et al., 2008)	**0.91**	15.90	8.75	10.90
	FCLS (Heinz and C.-I Chang, 2001)	**0.91**	14.27	5.48	9.73
FM	Bayesian (Halimi et al., 2011b)	13.09	**0.87**	7.87	9.72
	Taylor (Fan et al., 2009)	13.09	**0.86**	7.28	15.20
GBM	Bayesian (Halimi et al., 2011b)	1.87	10.10	5.25	9.34
	Taylor (Halimi et al., 2011b)	3.65	8.47	4.08	9.01
	Gradient (Halimi et al., 2011b)	2.47	2.45	**1.73**	8.69
PPNMM	Bayesian	1.58	1.98	**1.86**	**1.69**
	Taylor	**1.56**	2.21	1.88	**1.92**
	Gradient	1.69	1.98	1.98	**1.69**

Table 1.1 shows that the abundances estimated by the Bayesian algorithm and the LS methods are similar for the PPNMM. Moreover, for these 4 images, the PPNMM seems to be more robust than the other mixing models to deviations from the actual model.

The unmixing quality can also be evaluated by the average reconstruction error (ARE) defined by

$$\text{ARE} = \sqrt{\frac{1}{LN}\sum_{n=1}^{N}\|\hat{\mathbf{y}}_n - \mathbf{y}_n\|^2} \tag{1.37}$$

where $\mathbf{y}_n$ is the nth observed pixel and $\hat{\mathbf{y}}_n$ its estimate. Table 1.2 compares the AREs obtained for the different synthetic images. These results show that the AREs are close for the different unmixing algorithms even if the estimated abundances can vary more significantly. Again, the proposed PPNMM seems to be more robust than the other mixing models to deviations from the actual model in term of ARE.

Table 1.2: AREs ($\times 10^{-2}$): synthetic images .

		I_1 (LMM)	I_2 (FM)	I_3 (GBM)	I_4 (PPNMM)
LMM	Bayesian (Dobigeon et al., 2008)	**5.28**	6.54	5.65	5.89
	FCLS (Heinz and C.-I Chang, 2001)	**5.28**	5.74	5.42	5.48
FM	Bayesian	5.61	**5.29**	5.38	5.76
	Taylor (Fan et al., 2009)	5.61	**5.28**	5.38	5.75
GBM	Bayesian (Halimi et al., 2011b)	**5.29**	5.49	5.33	5.44
	Taylor (Halimi et al., 2011b)	5.31	5.40	5.30	5.42
	Gradient (Halimi et al., 2011b)	**5.29**	5.30	**5.28**	5.41
PPNMM	Bayesian	**5.28**	**5.29**	**5.28**	**5.28**
	Taylor	**5.29**	**5.29**	**5.28**	**5.28**
	Gradient	**5.29**	**5.29**	**5.28**	**5.28**

Fig. 1.2 shows the estimated distributions of b for the images $I_1, \ldots, I_4$ using the three presented algorithms (i.e., Bayesian, linearization and subgradient). This figure shows that the three algorithms perform similarly for the estimation of the nonlinearity parameter b.

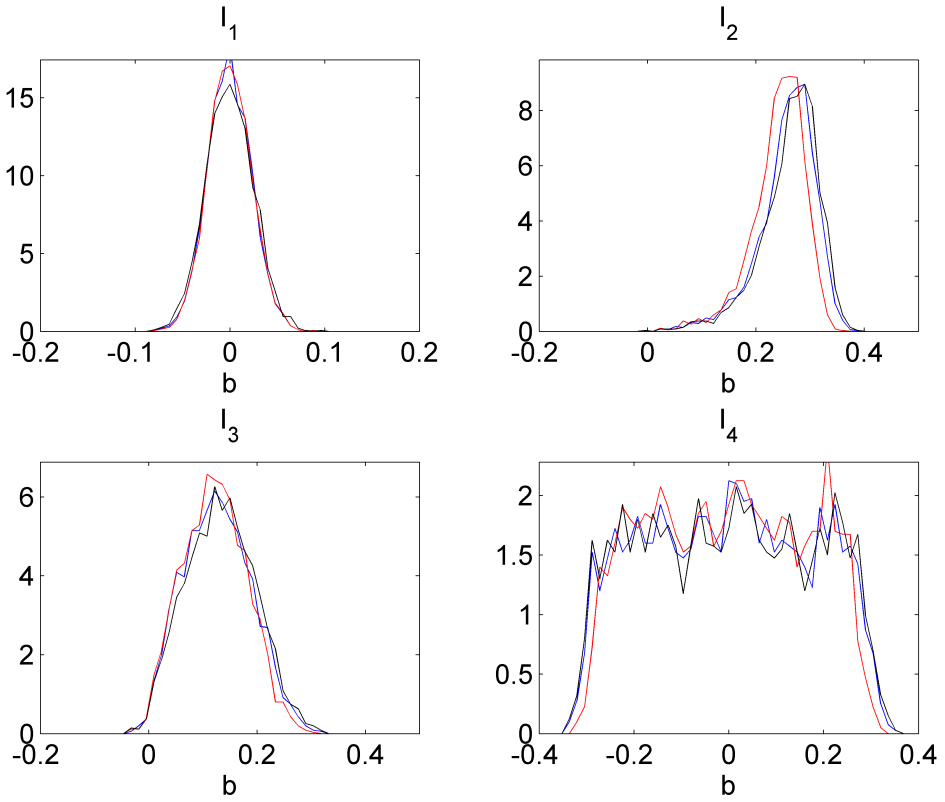


Figure 1.2: Histograms of the estimated nonlinearity parameter $\hat{b}$ for the four synthetic images estimated by the Bayesian (black), linearization-based (red) and subgradient-based (blue) algorithms.

Table 1.3 shows the execution times of MATLAB implementations on a 1.66GHz Dual-Core of the proposed algorithms for unmixing the proposed images (2500 pixels for each image). The linearization-based algorithm has the lowest computational cost and also provides accurate estimations. Note that the computational cost of the Bayesian algorithm (which allows prior knowledge to be included in the unmixing procedure) can be prohibitive for larger images and a high number of endmembers. However, the computational cost of the two proposed optimization methods (linearization and gradient-based) is very reasonable which make them very useful for practical applications.

Table 1.3: Computational times of the unmixing algorithms for 2500 pixels (in second).

	I_1	I_2	I_3	I_4
Bayesian	5960	6200	6600	5970
Taylor	5	10	8	7
Subgradient	84	102	96	101

The next set of simulations analyzes the performance of the proposed nonlinear SU algorithms for different numbers of endmembers ($R \in \{3, 6, 9, 12\}$) by unmixing 4 synthetic images of 500 pixels. The endmembers contained in these images have been randomly selected from the fourteen endmembers extracted by VCA from the full Cuprite scene described by Clark *et al.* (2003). For each image, the abundance vectors $\boldsymbol{a}_n, (n = 1, \ldots, 500)$ have been randomly generated according to a uniform distribution over the admissible set defined by the positivity and sum-to-one constraints. All images have been corrupted by an additive white Gaussian noise corresponding to a signal-to-noise ratio SNR = 20dB. The nonlinearity coefficients b are uniformly drawn in the set $(-0.3, 0.3)$. Table 1.4 compares the performance of the three proposed methods in term of abundance estimation and reconstruction error. These results show that the three methods perform similarly in term of reconstruction error. The Bayesian estimators tend to provide more accurate abundance estimations (i.e., smaller RNMSEs) for large values of R. Indeed, the Taylor and gradient algorithms may be trapped in local minima of the LS criterion (1.21) for large values of R.

Table 1.4: Unmixing performance of the supervised PPNMM-based algorithms for different R.

	Average RNMSEs($\times 10^{-2}$)				AREs($\times 10^{-2}$)			
	Bayesian		Taylor	Gradient	Bayesian		Taylor	Gradient
	MMSE	MAP			MMSE	MAP		
$R = 3$	7.50	10.42	9.43	9.41	4.18	4.22	4.17	4.17
$R = 6$	7.53	11.37	12.65	12.16	4.22	4.24	4.20	4.20
$R = 9$	5.69	9.56	11.90	11.41	4.27	4.29	4.24	4.24
$R = 12$	4.72	8.08	11.16	10.58	4.18	4.19	4.13	4.13

Real data

The first real image considered in this section is composed of $L = 189$ spectral bands and was acquired in 1997 by the airborne visible infrared imaging spectrometer (AVIRIS) over the Cuprite mining site in Nevada. A sub-image of size 50×50 pixels has been chosen here to evaluate the proposed unmixing procedures. The scene is mainly composed of muscovite, alunite and kaolinite, as explained by Dobigeon et al. (2009a). The endmembers extracted by VCA (Nascimento and Bioucas-Dias, 2005) and the nonlinear EEA proposed by Heylen et al. (2011) (referred to as "Heylen"), with $R = 3$ are depicted in Fig. 1.3. The endmembers obtained by the two methods have similar shapes. This result confirms the fact that the geometric EEAs (such as VCA) can be used as a first approximation for endmember estimation (Keshava and Mustard, 2002).

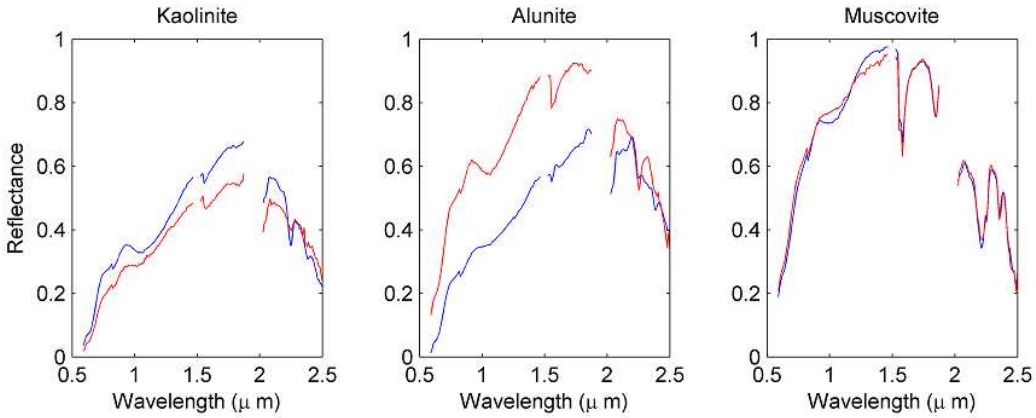


Figure 1.3: The $R = 3$ endmembers estimated by VCA (blue lines) and Heylen (red lines) for the Cuprite scene.

The three estimation algorithms presented above have been applied independently to each pixel of the scene using the endmembers extracted by the two EEAs. Examples of abundance maps obtained by the Heylen's method are presented in Fig. 1.4. They are similar to the abundance maps obtained with the FCLS algorithm which relies on the LMM.

However, the advantage of the PPNMM is that it allows the nonlinearities between the observations and the abundance vectors to be analyzed. For instance, Fig. 1.5 shows the estimated maps of b for the Cuprite image. These results show that the observations are nonlinearly related to the endmembers (since $b \neq 0$). However, the nonlinearity is weak since the estimated values of b are close to 0.

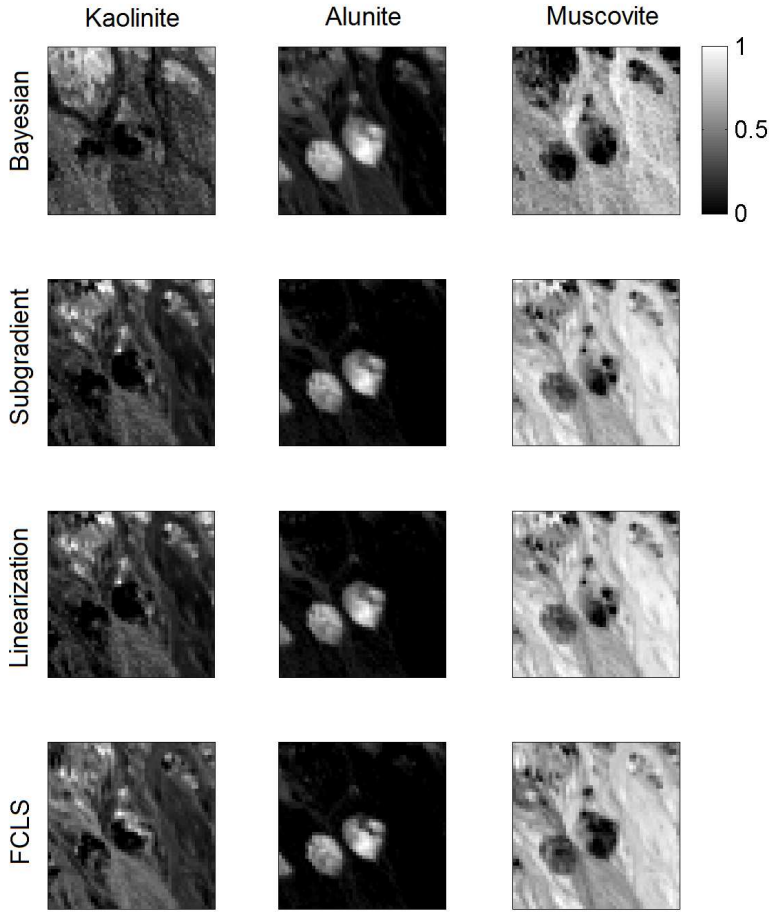


Figure 1.4: Abundance maps estimated by the Bayesian, linearization and subgradient methods for the Cuprite scene.

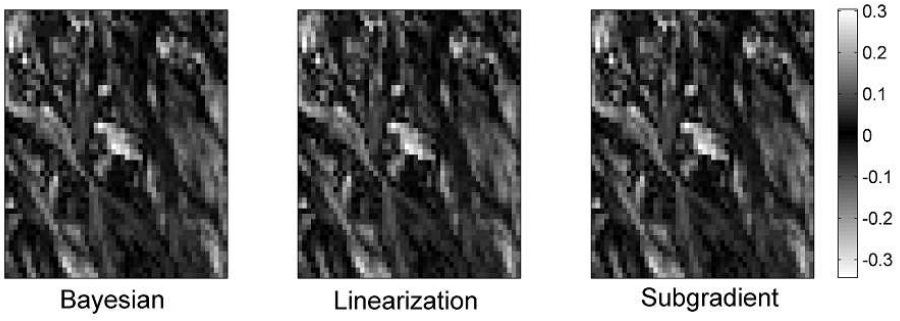


Figure 1.5: Maps of the nonlinearity parameter b estimated by the Bayesian, linearization and subgradient methods for the Cuprite scene.

The second real image considered in this section is composed of $L = 189$ spectral bands and was acquired in 1997 by the satellite AVIRIS over the Moffett Field, CA. A sub-image of size 50×50 pixels has also been chosen here to evaluate the proposed unmixing procedures. The scene is mainly composed of water, vegetation and soil. The endmembers extracted by VCA and Heylen's method with $R = 3$ are depicted in Fig. 1.6. Again, the endmembers

obtained by the two methods are similar. Examples of abundance maps estimated by the proposed algorithms with Heylen's method are presented in Fig. 1.7. They are similar to the abundance maps obtained with the FCLS algorithm which relies on the LMM. Fig. 1.8 shows the estimated maps of b for the Moffett image. In the water area, the observations are nonlinearly related to the endmembers (since $b \neq 0$). These nonlinearities can be due to the low amplitude of the water spectrum and nonlinear bathymetric effects.

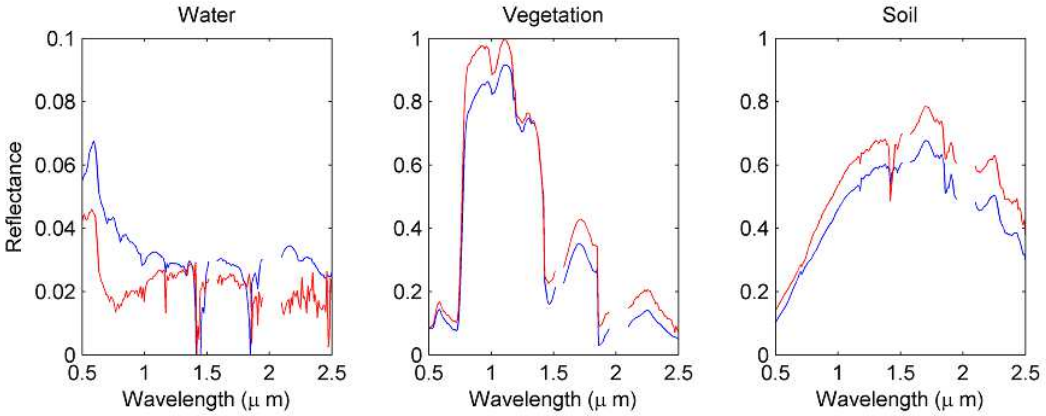


Figure 1.6: The $R = 3$ endmembers estimated by VCA (blue lines) and Heylen (red lines) for the Moffett scene.

The quality of unmixing is finally evaluated using the AREs for both real images. These AREs are compared in Table 1.5 with those obtained by assuming other mixing models. The proposed PPNMM provides smaller AREs when compared to other models which is a very encouraging result.

Table 1.5: AREs ($\times 10^{-2}$): Cuprite and Moffett images

		VCA		Heylen	
		Cuprite	Moffett	Cuprite	Moffett
LMM	Bayesian (Dobigeon et al., 2008)	2.14	2.70	2.35	2.02
	FCLS (Heinz and C.-I Chang, 2001)	2.11	2.62	2.10	2.00
FM	Bayesian	7.36	2.31	2.30	1.92
	Taylor (Fan et al., 2009)	3.05	2.29	2.29	1.92
GBM	Bayesian (Halimi et al., 2011b)	2.24	2.57	2.11	1.99
	Taylor (Halimi et al., 2011b)	2.34	2.41	2.03	2.01
	Gradient (Halimi et al., 2011b)	**2.02**	2.30	2.04	1.93
PPNMM	Bayesian	**1.19**	1.59	**1.91**	**1.85**
	Taylor	**1.19**	**1.54**	**1.90**	**1.84**
	Gradient	**1.19**	**1.55**	**1.90**	1.87

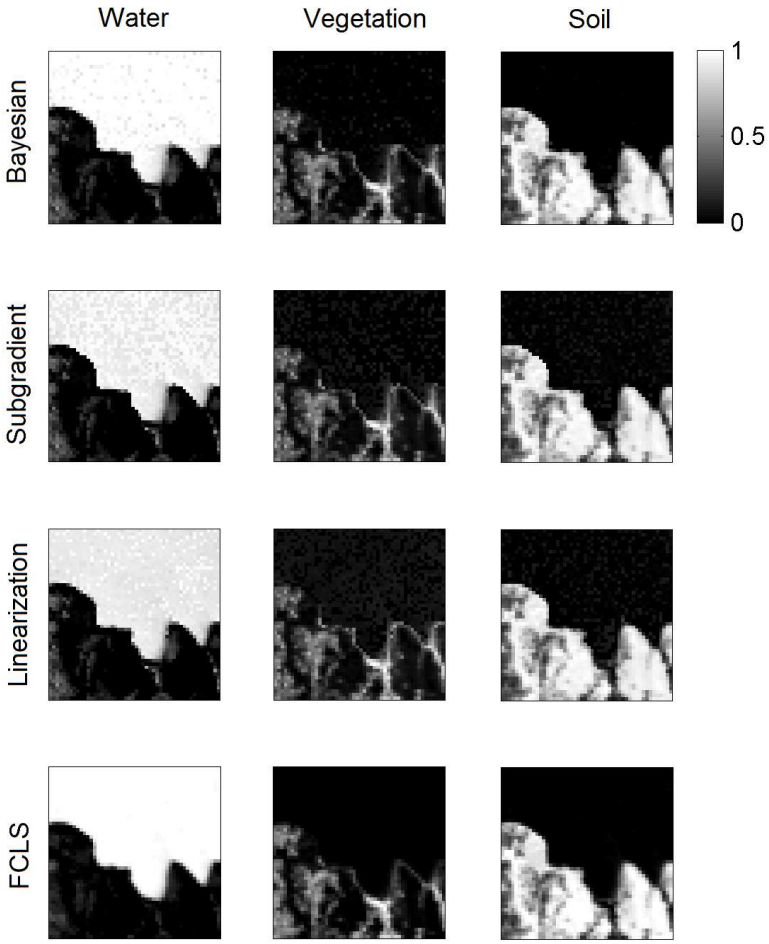


Figure 1.7: Abundance maps estimated by the Bayesian, linearization and subgradient methods for the Moffett scene.

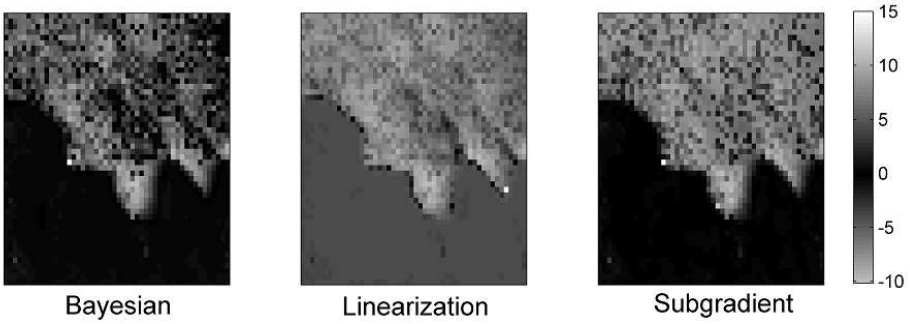


Figure 1.8: Maps of the nonlinearity parameter b estimated by the Bayesian, linearization and subgradient methods for the Moffett scene.

1.4.4 Intermediate conclusion

A Bayesian and two least squares algorithms were presented for supervised nonlinear SU of hyperspectral images. These algorithms assumed that the hyperspectral image pixels are related to the endmembers by a polynomial post-nonlinear mixing model. In the Bayesian framework, the constraints related to the unknown parameters were ensured by using appropriate prior distributions. The posterior distribution of the unknown parameter vector was then derived. The corresponding minimum mean square error estimator was approximated from samples generated using Markov chain Monte Carlo methods. Least squares methods were also investigated for unmixing the polynomial post-nonlinear model. These methods provided results similar to the Bayesian algorithm with a reduced computational cost, making them very attractive for hyperspectral image unmixing. Results obtained on synthetic and real images illustrated the accuracy of the polynomial post-nonlinear model and the performance of the corresponding estimation algorithms for supervised unmixing. In this study, we assumed that the endmembers were known (either coming from *a priori* or extracted from the data using a EEA). However, as it has been shown for linear SU, a joint estimation of the endmembers and mixing coefficients can provide more accurate mixture characterization, especially when there is not enough pure pixels in the observed image. This joint estimation is the aim of the last part of this chapter which addresses the problem of unsupervised SU using the PPNMM. The next section generalizes the Bayesian model proposed for supervised nonlinear unmixing to the case where the endmembers are unknown and to be estimated. Least-squares methods could also have been investigated. However, their convergence is difficult to prove and the proposed Bayesian algorithm provides accurate results in practice. Including the endmembers in the estimation procedure complicates the unmixing procedure and estimating more parameters usually requires a higher computational cost. To improve the mixing properties and the complexity of the sampler, Hamiltonian Monte Carlo methods are investigated.

1.5 Unsupervised PPNMM-based unmixing

In this second scenario, the spectral signatures of the endmembers contained in the hyperspectral image are unknown and thus to be estimated. Only the number of endmembers is assumed to be known. Consider a hyperspectral image consisting of N pixels distributed according to (1.3). The PPNMM defined in (1.3) allows the $L \times N$ observation matrix $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N]$ to be expressed as follows

$$\mathbf{Y} = \mathbf{MA} + [(\mathbf{MA}) \odot (\mathbf{MA})] \operatorname{diag}(\boldsymbol{b}) + \mathbf{E} \tag{1.38}$$

where $\mathbf{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N]$ is an $R \times N$ matrix, $\mathbf{E} = [\mathbf{e}_1, \ldots, \mathbf{e}_N]$ is an $L \times N$ matrix, $\boldsymbol{b} = [b_1, \ldots, b_N]^T$ is an $N \times 1$ vector containing the nonlinearity parameters. In this scenario, the noise sequences $\mathbf{e}_n, \forall n \in \{1, \ldots, N\}$ are additive independently distributed zero-mean Gaussian vectors with diagonal covariance matrix $\boldsymbol{\Sigma} = \operatorname{diag}\left(\boldsymbol{\sigma}^2\right)$, denoted as $\mathbf{e}_n \sim \mathcal{N}(\mathbf{0}_L, \boldsymbol{\Sigma})$, where $\boldsymbol{\sigma}^2 = [\sigma_1^2, \ldots, \sigma_L^2]^T$ is the vector of the L noise variances and $\operatorname{diag}\left(\boldsymbol{\sigma}^2\right)$ is an $L \times L$ diagonal matrix containing the elements of the vector $\boldsymbol{\sigma}^2$. Note that this noise characterization is more general than the one considered in the supervised SU scenario presented above. Precisely, the N noise vectors associated with the N pixels have noise variances differing from one spectral band to another, which is in agreement with real noise measurements. The abundance vectors in $\mathbf{A}$ satisfy the positivity and sum-to-one constraints (1.4) and the endmembers to be estimated are subject to the constraints (1.5).

1.5.1 Bayesian estimation

This section generalizes the hierarchical Bayesian model introduced in Section 1.4 in order to jointly estimate the abundances and endmembers, leading to a fully unsupervised hyperspectral unmixing algorithm. To handle abundance constraints, we propose to reparameterize the abundance vectors using the following transformation

$$a_{r,n} = \left(\prod_{k=1}^{r-1} z_{k,n}\right) \times \begin{cases} 1 - z_{r,n} & \text{if } r < R \\ 1 & \text{if } r = R \end{cases}. \tag{1.39}$$

This transformation has been recently suggested by Betancourt (2010). One motivation for using the latent variables $z_{r,n}$ instead of $a_{r,n}$ is the fact that the constraints (1.4) for the nth abundance vector $\boldsymbol{a}_n$ express as

$$0 < z_{r,n} < 1, \quad \forall r \in \{1, \ldots, R-1\} \tag{1.40}$$

for the nth coefficient vector $\boldsymbol{z}_n = [z_{1,n}, \ldots, z_{R-1,n}]^T$. As a consequence, the constraints (1.40) are much easier to handle for the sampling procedure than (1.4) (as will be shown in the next sections). The next section presents the Bayesian model associated with the unsupervised PPNMM (1.38) for SU.

Bayesian model

The unknown parameter vector associated with the PPNMM contains the reparameterized abundances $\mathbf{Z} = [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N]$ (satisfying the constraints (1.40)), the endmember matrix $\mathbf{M}$, the nonlinearity parameter vector $\boldsymbol{b}$ and the additive noise variance vector $\boldsymbol{\sigma}^2$. This section summarizes the likelihood and the parameter priors (associated with the proposed hierarchical Bayesian PPNMM) introduced to perform nonlinear unsupervised hyperspectral unmixing.

Likelihood Equation (1.38) shows that $\mathbf{y}_n|\mathbf{M}, \boldsymbol{z}_n, b_n, \boldsymbol{\sigma}^2$ is distributed according to a Gaussian distribution with mean $\mathbf{g}_n(\mathbf{M}\boldsymbol{a}_n)$ and covariance matrix $\boldsymbol{\Sigma}$, denoted as $\mathbf{y}_n|\mathbf{M}, \boldsymbol{z}_n, b_n, \boldsymbol{\sigma}^2 \sim \mathcal{N}(\mathbf{g}_n(\mathbf{M}\boldsymbol{a}_n), \boldsymbol{\Sigma})$. Note that the abundance vector $\boldsymbol{a}_n$ should be denoted as $\boldsymbol{a}_n(\boldsymbol{z}_n)$. However, the argument $\boldsymbol{z}_n$ has been omitted for brevity. Assuming independence between the observed pixels, the joint likelihood of the observation matrix $\mathbf{Y}$ can be expressed as

$$f(\mathbf{Y}|\mathbf{M}, \mathbf{Z}, \boldsymbol{b}, \boldsymbol{\sigma}^2) \propto |\boldsymbol{\Sigma}|^{-N/2} \text{etr}\left[-\frac{(\mathbf{Y}-\mathbf{X})^T\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\mathbf{X})}{2}\right] \tag{1.41}$$

where $\propto$ means "proportional to", $\text{etr}(\cdot)$ denotes the exponential trace and

$$\mathbf{X} = \mathbf{MA} + [(\mathbf{MA}) \odot (\mathbf{MA})]\,\text{diag}(\boldsymbol{b}) \tag{1.42}$$

is an $L \times N$ matrix.

Parameter priors To reflect the lack of prior knowledge about the abundances, we propose to assign prior distributions for the coefficient vector $\boldsymbol{z}_n$ that correspond to noninformative prior distributions for $\boldsymbol{a}_n$. More precisely, assigning the following beta priors

$$z_{n,r} \sim \mathcal{B}e(R-r, 1) \quad r \in \{1, \ldots, R-1\} \tag{1.43}$$

and assuming prior independence between the elements of $\boldsymbol{z}_n$ yield an abundance vector $\boldsymbol{a}_{\setminus R,n}$ uniformly distributed in the set defined in (1.10) (see (Betancourt, 2010) for details). Assuming prior independence between the coefficient vectors $\{\boldsymbol{z}_n\}_{n=1,\ldots,N}$ leads to

$$f(\mathbf{Z}) = \prod_{r=1}^{R-1}\left\{\frac{1}{B(R-r,1)^N}\prod_{n=1}^{N} z_{n,r}^{R-r-1}\right\} \tag{1.44}$$

where $B(\cdot,\cdot)$ is the Beta function.

For each endmember $\mathbf{m}_r$, we propose to use a Gaussian prior

$$\mathbf{m}_r \sim \mathcal{N}_{[0,1]^L}(\bar{\mathbf{m}}_r, s^2\mathbf{I}_L), \tag{1.45}$$

truncated on $[0,1]^L$ to satisfy the constraints (1.5). In this study, we propose to select the mean vectors $\bar{\mathbf{m}}_r$ as the pure components previously identified by the nonlinear EEA studied in (Heylen et al., 2011) and referred to as "Heylen". The variance s^2 reflects the degree of confidence given to this prior information. When no additional knowledge is available, this variance is fixed to a large value ($s^2 = 50$ in our simulations). Note that any EEA could be used to define the vectors $\bar{\mathbf{m}}_1, \ldots, \bar{\mathbf{m}}_R$. Without additional information, the unmixing problem assuming the PPNMM (1.38) may have an infinite set of solutions. However, considering the constraints (1.4) and (1.5) and enforcing $\mathbf{m}_r$ to be close to $\bar{\mathbf{m}}_r, \forall r$ removes this ambiguity.

It is interesting to note that the PPNMM reduces to the LMM for $b_n = 0$. Since the LMM is relevant for most observed pixels, it makes sense to assign prior distributions to the nonlinearity parameters that enforce sparsity for the vector $\boldsymbol{b}$. To detect linear and nonlinear mixtures of the pure spectral signatures in the image, the following conjugate Bernoulli-Gaussian prior is assigned to the nonlinearity parameter b_n

$$f(b_n|w, \sigma_b^2) = (1-w)\delta(b_n) + w\frac{1}{\sqrt{2\pi\sigma_b^2}}\exp\left(-\frac{b_n^2}{2\sigma_b^2}\right) \tag{1.46}$$

where $\delta(\cdot)$ denotes the Dirac delta function. Note that the prior distributions for the nonlinearity parameters $\{b_n\}_{n=1,\ldots,N}$ share the same hyperparameters $w \in [0,1]$ and $\sigma_b^2 \in \mathbb{R}^+$. More precisely, the weight w is the prior probability of having a nonlinearly mixed pixel in the image. Assuming prior independence between the nonlinearity parameters $\{b_n\}_{n=1,\ldots,N}$, the joint prior distribution of the nonlinearity parameter vector $\boldsymbol{b}$ can be expressed as follows

$$f(\boldsymbol{b}|w,\sigma_b^2) = \prod_{n=1}^{N} f(b_n|w,\sigma_b^2). \tag{1.47}$$

A Jeffreys' prior is chosen for the noise variance of each spectral band σ_ℓ^2

$$f(\sigma_\ell^2) \propto \frac{1}{\sigma_\ell^2}\mathbf{1}_{\mathbb{R}^+}\left(\sigma_\ell^2\right) \tag{1.48}$$

which reflects the absence of knowledge for this parameter (see (Bernardo and Smith, 1994) for motivations). Assuming prior independence between the noise variances, we obtain

$$f(\boldsymbol{\sigma}^2) = \prod_{\ell=1}^{L} f(\sigma_\ell^2). \tag{1.49}$$

Hyperparameter priors The performance of the proposed Bayesian model for spectral unmixing depends on the values of the hyperparameters σ_b^2 and w. When the hyperparameters are difficult to adjust, it is classical to include them in the unknown parameter vector, resulting in a hierarchical Bayesian model (Altmann et al., 2012; Dobigeon et al., 2009a; Robert, 2007). This strategy requires to define prior distributions for the hyperparameters. A conjugate inverse-Gamma prior is assigned to σ_b^2

$$\sigma_b^2 \sim \mathcal{IG}\left(\gamma,\nu\right) \tag{1.50}$$

where (γ,ν) are real parameters fixed to obtain a flat prior, reflecting the absence of knowledge about the variance σ_b^2 ((γ,ν) will be set to $(10^{-1},10^{-1})$ in the simulation section). A uniform prior distribution is assigned to the hyperparameter w

$$w \sim \mathcal{U}_{[0,1]}(w) \tag{1.51}$$

since there is no a priori information regarding the proportions of linearly and nonlinearly mixed pixels in the image. The resulting directed acyclic graph (DAG) associated with the proposed Bayesian model is depicted in Fig. 1.9.

Joint posterior distribution The joint posterior distribution of the unknown parameters $\{\boldsymbol{\theta},\boldsymbol{\Phi}\}$ where $\boldsymbol{\theta} = \left\{\mathbf{Z},\mathbf{M},\boldsymbol{b},\boldsymbol{\sigma}^2\right\}$ and $\boldsymbol{\Phi} = \left\{\sigma_b^2, w\right\}$ can be computed using the following hierarchical structure

$$f(\boldsymbol{\theta},\boldsymbol{\Phi}|\mathbf{Y}) \propto f(\mathbf{Y}|\boldsymbol{\theta},\boldsymbol{\Phi})f(\boldsymbol{\theta},\boldsymbol{\Phi}) \tag{1.52}$$

where $f(\mathbf{Y}|\boldsymbol{\theta})$ has been defined in (4.3). By assuming *a priori* independence between the parameters $\mathbf{Z}$, $\mathbf{M}$, $\boldsymbol{b}$ and $\boldsymbol{\sigma}^2$ and between the hyperparameters σ_b^2 and w, the joint prior distribution of the unknown parameter vector can be expressed as

$$\begin{aligned} f(\boldsymbol{\theta},\boldsymbol{\Phi}) &= f(\boldsymbol{\theta}|\boldsymbol{\Phi})f(\boldsymbol{\Phi}) \\ &= f(\mathbf{Z})f(\mathbf{M})f(\boldsymbol{\sigma}^2)f(\boldsymbol{b}|\sigma_b^2,w)f(\sigma_b^2)f(w). \end{aligned} \tag{1.53}$$

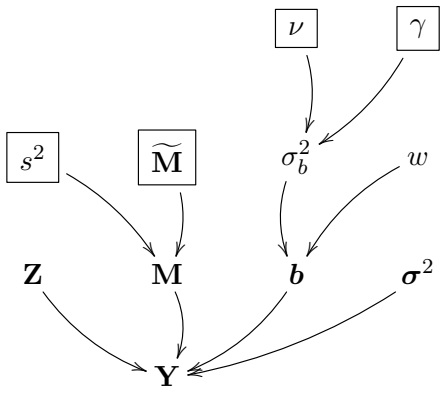


Figure 1.9: DAG for the parameter and hyperparameter priors (the fixed parameters appear in boxes).

The joint posterior distribution $f(\boldsymbol{\theta}, \boldsymbol{\Phi}|\mathbf{Y})$ can then be computed up to a multiplicative constant after replacing (1.53) and (4.3) in (4.17). Unfortunately, it is difficult to obtain closed form expressions for the standard Bayesian estimators (including the maximum a posteriori (MAP) and the minimum mean square error (MMSE) estimators) associated with (4.17). In this study, we propose to use efficient Markov Chain Monte Carlo (MCMC) methods to generate samples asymptotically distributed according to (4.17). Due to the large number of parameters to be sampled, we use an HMC algorithm which allows the number of sampling steps to be reduced and which improves the mixing properties of the sampler. The generated samples are then used to compute the MMSE estimator of the unknown parameter vector $(\boldsymbol{\theta}, \boldsymbol{\Phi})$. The next section summarizes the basic principles of the HMC methods that will be used to sample asymptotically from (4.17).

Constrained Hamiltonian Monte Carlo method

HMCs are powerful methods for sampling continuous distributions by introducing fictitious momentum variables. Let $\mathbf{q} \in \mathbb{R}^D$ be the parameter of interest and $\pi(\mathbf{q})$ its corresponding distribution to be sampled from. From statistical mechanics, the distribution $\pi(\mathbf{q})$ can be related to a potential energy function $U(\mathbf{q}) = -\log[\pi(\mathbf{q})] + c$ where c is a positive constant such that $\int \exp(-U(\mathbf{q}) + c)\, \mathrm{d}\mathbf{q} = 1$. The Hamiltonian of $\pi(\mathbf{q})$ is a function of the energy $U(\mathbf{q})$ and of an additional momentum vector $\mathbf{p} \in \mathbb{R}^D$ defined as

$$H(\mathbf{q}, \mathbf{p}) = U(\mathbf{q}) + K(\mathbf{p}) \tag{1.54}$$

where $K(\mathbf{p})$ is an arbitrary kinetic energy function. Usually, a quadratic kinetic energy is chosen and we propose to use $K(\mathbf{p}) = \mathbf{p}^T\mathbf{p}/2$ in this study (for reasons explained later). The Hamiltonian (1.54) defines the following distribution for $(\mathbf{q}, \mathbf{p})$

$$\begin{aligned} f(\mathbf{q}, \mathbf{p}) &\propto \exp[-H(\mathbf{q}, \mathbf{p})] \\ &\propto \pi(\mathbf{q}) \exp\left(-\frac{1}{2}\mathbf{p}^T\mathbf{p}\right) \end{aligned} \tag{1.55}$$

which shows that $\mathbf{q}$ and $\mathbf{p}$ are independent and that the marginal distribution of $\mathbf{p}$ is the $\mathcal{N}(\mathbf{0}_D, \mathbf{I}_D)$ distribution. The HMC algorithm allows samples to be asymptotically generated according to (1.55). The ith HMC iteration uses the pair of vectors $(\mathbf{q}^{(i)}, \mathbf{p}^{(i)})$ and consists of two steps. The first step resamples the initial momentum $\tilde{\mathbf{p}}^{(i)}$ according to the standard multivariate Gaussian distribution. The second step uses Hamiltonian dynamics to

propose a candidate $(\mathbf{q}^*, \mathbf{p}^*)$ which is accepted with the following probability

$$\rho = \min\left\{\exp\left[-H(\mathbf{q}^*, \mathbf{p}^*) + H(\mathbf{q}^{(i)}, \tilde{\mathbf{p}}^{(i)})\right], 1\right\}. \tag{1.56}$$

Generation of the candidate $(\mathbf{q}^*, \mathbf{p}^*)$: Hamiltonian dynamics are usually simulated by discretization methods such as Euler or leapfrog methods. The classical leapfrog method is a discretization scheme composed of N_{LF} steps with a discretization stepsize ϵ. The nth leapfrog step can be expressed as

$$\mathbf{p}^{(i,n\epsilon+\epsilon/2)} = \mathbf{p}^{(i,n\epsilon)} - \frac{\epsilon}{2}\frac{\partial U}{\partial \mathbf{q}^T}\left(\mathbf{q}^{(i,n\epsilon)}\right) \tag{1.57a}$$

$$\mathbf{q}^{(i,(n+1)\epsilon)} = \mathbf{q}^{(i,n\epsilon)} + \epsilon\mathbf{p}^{(i,n\epsilon+\epsilon/2)} \tag{1.57b}$$

$$\mathbf{p}^{(i,(n+1)\epsilon)} = \mathbf{p}^{(i,n\epsilon+\epsilon/2)} - \frac{\epsilon}{2}\frac{\partial U}{\partial \mathbf{q}^T}\left[\mathbf{q}^{(i,(n+1)\epsilon)}\right]. \tag{1.57c}$$

The leapfrog method starts with $(\mathbf{q}^{(i,0)}, \tilde{\mathbf{p}}^{(i)}) = (\mathbf{q}^{(i)}, \tilde{\mathbf{p}}^{(i)})$ and the candidate is set after N_{LF} steps to $(\mathbf{q}^*, \mathbf{p}^*) = (\mathbf{q}^{(i,\epsilon N_{\text{LF}})}, \tilde{\mathbf{p}}^{(i,\epsilon N_{\text{LF}})})$.

However, if $\mathbf{q}$ is subject to constraints, more sophisticated discretization methods must be used. Assume that the vector of interest $\mathbf{q} = [q_1, \ldots, q_D]^T$ satisfies the following constraints

$$q_l < q_d < q_u, \quad d \in \{1, \ldots, D\} \tag{1.58}$$

where q_l (resp. q_u) is the lower (resp. upper) bound for q_d (such kind of constraints need to be satisfied by the elements of $\mathbf{Z}$ and the endmembers in $\mathbf{M}$). In this study we propose to use the constrained leapfrog scheme studied in (Brooks, 2011, Chap. 5), consisting of N_{LF} steps, with a discretization stepsize ϵ_q. Each constrained HMC (CHMC) iteration starts in a similar way to the classical leapfrog method, with the sequential sampling of the momentum $\mathbf{p}$ (1.57a) and the vector $\mathbf{q}$ (1.57b). However, if the generated vector $\mathbf{q}$ violates the constraints (1.58), it is modified depending on the violated constraints and the momentum is negated (see (Brooks, 2011, Chap. 5) for more details). This step is repeated until each component of the generated $\mathbf{q}$ satisfies the constraints. The CHMC ends with the update of the momentum $\mathbf{p}$ (1.57c). One iteration of the resulting constrained HMC algorithm (CHMC) is summarized in Algo. 1.4. As mentioned above, one might think of using a more sophisticated kinetic energy for $\mathbf{p}$ to improve the performance of the HMC algorithm. However, the kinetic energy $K(\mathbf{p}) = \mathbf{p}^T\mathbf{p}/2$ allows the discretization method handling the constraints to be simple and will provide good performance for our application (as will be shown in Section 4.7). The performance of the HMC mainly relies on the values of the parameters N_{LF} and ϵ_q. Fortunately, the choice of ϵ_q is almost independent of N_{LF} such that these two parameters can be tuned sequentially. The procedures used in this study to adjust N_{LF} and ϵ_q are detailed in the next paragraphs.

%Initialization of the ith iteration($n = 0$)

- $\mathbf{q}^{(i,0)} = \mathbf{q}^{(i)}$ satisfying the constraints (1.58)
- Sample $\mathbf{p}^{(i,0)} = \tilde{\mathbf{p}}^{(i)}) \sim \mathcal{N}(\mathbf{0}_D, \mathbf{I}_D)$

%Modified leapfrog steps
for $n = 0 : N_{\text{LF}} - 1$ **do**
  %Standard leapfrog steps
  - Compute $\mathbf{p}^{(i,n\epsilon+\epsilon/2)} = \mathbf{p}^{(i,n\epsilon)} - \frac{\epsilon}{2}\frac{\partial U}{\partial \mathbf{q}^T}\left(\mathbf{q}^{(i,n\epsilon)}\right)$
  - Compute $\mathbf{q}^{(i,(n+1)\epsilon)} = \mathbf{q}^{(i,n\epsilon)} + \epsilon\mathbf{p}^{(i,n\epsilon+\epsilon/2)}$

  %Steps required to ensure $\mathbf{q}^{(i,(n+1)\epsilon)}$ satisfies (1.58)
  while $\mathbf{q}^{(i,(n+1)\epsilon)}$ does not satisfy (1.58) **do**
    for $d = 1 : D$ **do**
      if $q_d^{(i,(n+1)\epsilon)} < q_l$ **then**
        Set $q_d^{(i,(n+1)\epsilon)} = 2q_l - q_d^{(i,(n+1)\epsilon)}$
        (replace $q_d^{(i,(n+1)\epsilon)}$ by its symmetric with respect to q_l)
        Set $p_d^{(i,n\epsilon+\epsilon/2)} = -p_d^{(i,n\epsilon+\epsilon/2)}$
      end if
      if $q_d^{(t+\epsilon)} > q_u$ **then**
        Set $q_d^{(i,(n+1)\epsilon)} = 2q_u - q_d^{(i,(n+1)\epsilon)}$
        (replace $q_d^{(i,(n+1)\epsilon)}$ by its symmetric with respect to q_u)
        Set $p_d^{(i,n\epsilon+\epsilon/2)} = -p_d^{(i,n\epsilon+\epsilon/2)}$
      end if
    end for
  end while
  %Standard leapfrog step
  Compute $\mathbf{p}^{(i,(n+1)\epsilon)} = \mathbf{p}^{(i,n\epsilon+\epsilon/2)} - \frac{\epsilon}{2}\frac{\partial U}{\partial \mathbf{q}^T}\left[\mathbf{q}^{(i,(n+1)\epsilon)}\right]$
end for
%Accept-reject procedure
Set $\mathbf{p}^* = \mathbf{p}^{(i,\epsilon N_{\text{LF}})}$ and $\mathbf{q}^* = \mathbf{q}^{(i,\epsilon N_{\text{LF}})}$
Compute ρ using (1.56)
Set $(\mathbf{q}^{(i+1)}, \mathbf{p}^{(i+1)}) = (\mathbf{q}^*, \mathbf{p}^*)$ with probability ρ
Else set $(\mathbf{q}^{(i+1)}, \mathbf{p}^{(i+1)}) = (\mathbf{q}^{(i)}, \tilde{\mathbf{p}}^{(i)})$.

ALGO. 1.4: Constrained Hamiltonian Monte Carlo iteration

Tuning the stepsize ϵ_q: The step size ϵ_q is related to the accuracy of the leapfrog method to approximate the Hamiltonian dynamics. When ϵ_q is "small", the approximation of the Hamiltonian dynamic is accurate and the acceptance rate (1.56) is high. However, the exploration of the distribution support is slow (for a given N_{LF}). In this study, we propose to tune the stepsize during the burn-in period of the sampler. More precisely, the stepsize is decreased (resp. increased) by 25% if the average acceptance rate over the last 50 iterations is smaller than 0.5 (resp. higher than 0.8). Note that the stepsize update only happens during the burn-in period to ensure the Markov

chain is homogeneous after the burn-in period.

Tuning the number of leapfrog steps N_{LF}: Assume ϵ_q has been correctly adjusted. Too small values of N_{LF} lead to a slow exploration of the distribution (random walk behavior) whereas too high values of N_{LF} require high computational time. Similarly to the stepsize ϵ_q, the optimal choice of N_{LF} depends on the distribution to be sampled. The sampling procedure proposed in this study consists of several HMC updates included in a Gibbs sampler (as will be shown in the next section). The number of leapfrog steps required for each of these CHMC updates has been adjusted by cross-validation. From preliminary runs, we have observed that setting the number of leapfrog steps for each HMC update close to $N_{\mathrm{LF}} = 50$ provides a reasonable tradeoff ensuring a good exploration of the target distribution and a reasonable computational complexity. To avoid possible periodic trajectories, it is recommended to let N_{LF} random (Brooks, 2011, Chap. 5). In this study, we have assumed that N_{LF} is uniformly drawn in the interval $[45, 55]$ at each iteration of the Gibbs sampler. The next section presents the Gibbs sampler (including CHMC steps) which is proposed to sample according to (4.17).

Gibbs Sampler

The principle of the Gibbs sampler is to sample according to the conditional distributions of the posterior of interest (Robert and Casella, 2004, Chap. 10). Due to the large number of parameters to be estimated, it makes sense to use a block Gibbs sampler to improve the convergence of the sampling procedure. More precisely, we propose to sample sequentially $\mathbf{M}, \mathbf{Z}, \boldsymbol{b}, \boldsymbol{\sigma}^2, \sigma_b^2$ and w using six moves that are detailed in the next sections.

Sampling the coefficient matrix Z Sampling from $f(\mathbf{Z}|\mathbf{Y}, \mathbf{M}, \boldsymbol{b}, \boldsymbol{\sigma}^2, \sigma_b^2, w)$ is difficult due to the complexity of this distribution. In this case, it is classical to use an accept/reject procedure to update the coefficient matrix $\mathbf{Z}$ (hybrid Metropolis-Within-Gibbs sampler). Since the elements of $\mathbf{Z}$ satisfy the constraints (1.40), the CHMC studied in Section 1.5.1 could be used to sample according to the conditional distribution $f(\mathbf{Z}|\mathbf{Y}, \mathbf{M}, \boldsymbol{b}, \boldsymbol{\sigma}^2, \sigma_b, w)$. However, as for Metropolis-Hastings updates, the convergence of HMCs generally slows down when the dimensionality of the vector to be sampled increases. Consequently, sampling an $N(R-1)$-dimensional vector using the proposed CHMC can be inefficient when the number of pixels is very large. It can be shown that

$$f(\mathbf{Z}|\mathbf{Y}, \mathbf{M}, \boldsymbol{b}, \boldsymbol{\sigma}^2, \sigma_b, w) = \prod_{n=1}^{N} f(\boldsymbol{z}_n|\mathbf{y}_n, \mathbf{M}, b_n, \boldsymbol{\sigma}^2), \tag{1.59}$$

i.e., the N coefficients vectors $\{\boldsymbol{z}_n\}_{n=1,\ldots,N}$ are a posteriori independent and can be sampled independently in a parallel manner. Moreover, straightforward computations lead to

$$f(\boldsymbol{z}_n|\mathbf{y}_n, \mathbf{M}, b_n, \boldsymbol{\sigma}^2) \propto \exp\left(-\frac{(\mathbf{y}_n - \mathbf{x}_n)^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}_n - \mathbf{x}_n)}{2}\right) \mathbf{1}_{(0,1)^{R-1}}(\boldsymbol{z}_n) \left(\prod_{r}^{R-1} z_{n,r}^{R-r-1}\right) \tag{1.60}$$

where $\mathbf{x}_n = \mathbf{g}_n(\mathbf{M}\boldsymbol{a}_n)$, $\mathbf{1}_{(0,1)^{R-1}}(\cdot)$ denotes the indicator function over $(0,1)^{R-1}$. The distribution (1.60) is related to the following potential energy

$$U(\boldsymbol{z}_n) = \frac{(\mathbf{y}_n - \mathbf{x}_n)^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}_n - \mathbf{x}_n)}{2} - \sum_{r=1}^{R-1} \log\left(z_{n,r}^{R-r-1}\right) \tag{1.61}$$

where we note that $f(\boldsymbol{z}_n|\mathbf{y}_n, \mathbf{M}, b_n, \boldsymbol{\sigma}^2) \propto \exp[-U(\boldsymbol{z}_n)]$. N momentum vectors associated with a canonical kinetic energy are introduced. The CHMC of Section 1.5.1 is then applied independently to the N vectors $\boldsymbol{z}_n$ whose

dimension $(R-1)$ is relatively small. The partial derivatives of the potential function (1.61) required in Algo. 1.4 are derived in Appendix C.

Sampling the endmember matrix M From (4.17) and (1.53), it can be seen that

$$f(\mathbf{M}|\mathbf{Y},\mathbf{Z},\boldsymbol{b},\boldsymbol{\sigma}^2,s^2,\widetilde{\mathbf{M}}) = \prod_{\ell=1}^{L} f(\mathbf{m}_{\ell,:}|\mathbf{y}_{\ell,:},\mathbf{Z},\boldsymbol{b},\sigma_\ell^2,s^2,\bar{\mathbf{m}}_{\ell,:})$$

where $\mathbf{m}_{\ell,:}$ (resp. $\bar{\mathbf{m}}_{\ell,:}$ and $\mathbf{y}_{\ell,:}$) is the ℓth row of $\mathbf{M}$ (resp. of $\widetilde{\mathbf{M}}$ and $\mathbf{Y}$) and

$$f(\mathbf{m}_{\ell,:}|\mathbf{y}_{\ell,:},\mathbf{Z},\boldsymbol{b},\sigma_\ell^2,s^2,\bar{\mathbf{m}}_{\ell,:}) \propto \exp\left(-\frac{\|\mathbf{y}_{\ell,:}-\mathbf{t}_\ell\|^2}{2\sigma_\ell^2}\right)\exp\left(-\frac{\|\mathbf{m}_{\ell,:}-\bar{\mathbf{m}}_{\ell,:}\|^2}{2s^2}\right)\mathbf{1}_{(0,1)^R}(\mathbf{m}_{\ell,:}) \tag{1.62}$$

with $\mathbf{t}_\ell = \mathbf{A}^T\mathbf{m}_{\ell,:} + \operatorname{diag}(\boldsymbol{b})\left[\left(\mathbf{A}^T\mathbf{m}_{\ell,:}\right)\odot\left(\mathbf{A}^T\mathbf{m}_{\ell,:}\right)\right]$. Consequently, the rows of the endmember matrix $\mathbf{M}$ can be sampled independently similarly to the procedure described in the previous section (to sample $\mathbf{Z}$). More precisely, we introduce a potential energy $V(\mathbf{m}_{\ell,:})$ associated with $\mathbf{m}_{\ell,:}$ defined by

$$V(\mathbf{m}_{\ell,:}) \quad = \quad \frac{\|\mathbf{y}_{\ell,:}-\mathbf{t}_\ell\|^2}{2\sigma_\ell^2} + \frac{\|\mathbf{m}_{\ell,:}-\bar{\mathbf{m}}_{\ell,:}\|^2}{2s^2} \tag{1.63}$$

and a momentum vector associated with a canonical kinetic energy. The partial derivatives of the potential function (1.63) required in Algo. 1.4 are derived in Appendix C.

Sampling the nonlinearity parameter vector b Using (4.17) and (1.53), it can be easily shown that the conditional distribution of $b_n|\mathbf{y}_n,\mathbf{M}\boldsymbol{z}_n,\boldsymbol{\sigma}^2,w,\sigma_b^2$ is the following Bernoulli-Gaussian distribution

$$b_n|\mathbf{y}_n,\mathbf{M},\boldsymbol{z}_n,\boldsymbol{\sigma}^2,w,\sigma_b^2 \sim (1-w_n^*)\delta(b_n) + w_n^*\mathcal{N}\left(\mu_n,s_n^2\right) \tag{1.64}$$

where

$$\mu_n = \frac{\sigma_b^2\left(\mathbf{y}_n-\mathbf{M}\boldsymbol{a}_n\right)^T\boldsymbol{\Sigma}^{-1}\mathbf{h}_n}{\sigma_b^2\mathbf{h}_n^T\boldsymbol{\Sigma}^{-1}\mathbf{h}_n+1}, \quad s_n^2 = \frac{\sigma_b^2}{\sigma_b^2\mathbf{h}_n^T\boldsymbol{\Sigma}^{-1}\mathbf{h}_n+1}$$

and $\mathbf{h}_n = (\mathbf{M}\boldsymbol{a}_n)\odot(\mathbf{M}\boldsymbol{a}_n)$. Moreover,

$$\begin{aligned} w_n^* &= \frac{w}{\beta_n + w(1-\beta_n)} \\ \beta_n &= \frac{\sigma_b}{s_n}\exp\left(-\frac{\mu_n^2}{2s_n^2}\right). \end{aligned} \tag{1.65}$$

For each b_n, the conditional distribution (1.64) does not depend on $\{b_k\}_{k\neq n}$. Consequently, the nonlinearity parameters $\{b_n\}_{n=1,\ldots,N}$ can be sampled independently in a parallel manner.

Sampling the noise variance vector $\boldsymbol{\sigma}^2$ By considering the posterior distribution (4.17), it can be shown that

$$f(\boldsymbol{\sigma}^2|\mathbf{Y},\mathbf{M},\mathbf{Z},\boldsymbol{b}) = \prod_{\ell=1}^{L} f(\sigma_\ell^2|\mathbf{y}_{\ell,:},\mathbf{m}_{:,\ell},\mathbf{Z},\boldsymbol{b}) \tag{1.66}$$

and that $\sigma_\ell^2|\mathbf{y}_{\ell,:},\mathbf{m}_{:,\ell},\mathbf{Z},\boldsymbol{b}$ is distributed according to the following inverse-gamma distribution

$$\sigma_\ell^2|\mathbf{y}_{\ell,:},\mathbf{m}_{:,\ell},\mathbf{Z},\boldsymbol{b} \sim \mathcal{IG}\left(\frac{N}{2},\frac{(\mathbf{y}_{\ell,:}-\mathbf{x}_{\ell,:})^T(\mathbf{y}_{\ell,:}-\mathbf{x}_{\ell,:})}{2}\right) \tag{1.67}$$

where $\mathbf{X} = [\mathbf{x}_{1,:},\ldots,\mathbf{x}_{L,:}]^T$. Thus the noise variances can be sampled easily and independently.

1: Initialization $t = 0$

- $\mathbf{Z}^{(0)}, \mathbf{M}^{(0)}, \boldsymbol{b}^{(0)}, \boldsymbol{\sigma}^{2(0)}, w^{(0)}, \sigma_b^{2(0)}$.

2: Iterations
3: **for** $t = 1 : N_{\text{MC}}$ **do**
4: Parameter update
5: Sample $\mathbf{Z}^{(t)}$ from the pdfs (1.60) using a CHMC procedure.
6: Sample $\mathbf{M}^{(t)}$ from the pdfs (1.62) using a CHMC procedure.
7: Sample $\boldsymbol{b}^{(t)}$ from the pdfs (1.64).
8: Sample $\boldsymbol{\sigma}^{2(t)}$ from the pdfs (1.67).
9: Hyperparameter update
10: Sample $\sigma_b^{2(t)}$ from the pdf (1.68).
11: Sample $w^{(t)}$ from the pdf (1.69).
12: **end for**

ALGO. 1.5: Gibbs Sampling Algorithm

Sampling the hyperparameters σ_b^2 and w Looking carefully at the posterior distribution (4.17), it can be seen that $\sigma_b^2|\boldsymbol{b}, \gamma, \nu$ is distributed according to the following inverse-gamma distribution

$$\sigma_b^2|\boldsymbol{b}, \gamma, \nu \sim \mathcal{IG}\left(\frac{n_1}{2} + \gamma, \sum_{n \in I_1} \frac{b_n^2}{2} + \nu\right) \tag{1.68}$$

with $I_1 = \{n|b_n \neq 0\}$, $n_0 = \|\boldsymbol{b}\|_0$ (where $\|\cdot\|_0$ is the ℓ_0 norm, i.e., the number of elements of $\boldsymbol{b}$ that are different from zero) and $n_1 = N - n_0$, from which it is easy to sample. Similarly, we obtain

$$w|\boldsymbol{b} \sim \mathcal{B}e(n_1 + 1, n_0 + 1). \tag{1.69}$$

Finally, the Gibbs sampler (including HMC procedures) used to sample according to the posterior (4.17) consists of the six steps summarized in Algo. 1.5. The small number of sampling steps is due to the high parallelization properties of the proposed sampling procedure, i.e., the generation of the N coefficient vectors $\{\boldsymbol{z}_n\}_{n=1,\ldots,N}$, the N nonlinearity parameters $\{b_n\}_{n=1,\ldots,N}$ and the L reflectance vectors $\{\mathbf{m}_{\ell,:}\}_{\ell=1,\ldots,L}$. After generating N_{MC} samples using the procedures detailed above, the MMSE estimator of the unknown parameters can be approximated by computing the empirical averages of these samples, after an appropriate burn-in period[4]. The next section studies the performance of the proposed algorithm for synthetic hyperspectral images.

1.5.2 Simulations

Simulations on synthetic data

The performance of the proposed nonlinear SU algorithm is first evaluated by unmixing 3 synthetic images of size 50×50 pixels. The $R = 3$ endmembers observed at $L = 207$ different spectral bands and contained in these

[4]The length of the burn-in period has been determined using appropriate convergence diagnoses (Robert and Cellier, 1998).

images have been extracted from the spectral libraries provided with the ENVI software (RSI (Research Systems Inc.), 2003) (i.e., green grass, olive green paint and galvanized steel metal). The first synthetic image I_1 has been generated using the standard linear mixing model (LMM). A second image I_2 has been generated according to the PPNMM and a third image I_3 has been generated according to the generalized bilinear mixing model (GBM) presented in (Halimi et al., 2011a). For each image, the abundance vectors $\boldsymbol{a}_n, n = 1, \ldots, 2500$ have been randomly generated according to a uniform distribution in the admissible set defined by

$$\mathcal{S}_{\mathrm{t}} = \left\{ \boldsymbol{a} \left| 0 < a_r < 0.9, \sum_{r=1}^{R} a_r = 1 \right. \right\}. \tag{1.70}$$

Note that the conditions $a_r < 0.9$ ensure that there is no pure pixel in the images, which makes the unmixing problem more challenging. All images have been corrupted by an additive independent and identically distributed (i.i.d) Gaussian noise of variance $\sigma^2 = 10^{-4}$, corresponding to an average signal-to-noise ratio SNR $\simeq$ 21dB for the three images. The noise is assumed to be i.i.d. to fairly compare unmixing performance with SU algorithms assuming i.i.d. Gaussian noise. The nonlinearity coefficients are uniformly drawn in the set $[0, 1]$ for the GBM. The parameters $b_n, n = 1, \ldots, N$ have been generated uniformly in the set $[-0.3, 0.3]$ for the PPNMM.

Table 1.6: Abundance RNMSEs ($\times 10^{-2}$): synthetic images.

		I_1 (LMM)	I_2 (PPNMM)	I_3 (GBM)
LMM	SLMM	3.78	13.21	6.83
	ULMM	0.66	10.87	4.21
PPNMM	SPPNMM	4.18	6.04	4.13
	UPPNMM	**0.37**	**0.81**	**1.38**
GBM		4.18	11.15	5.02

Different estimation procedures have been considered for the three mixing models. More precisely,

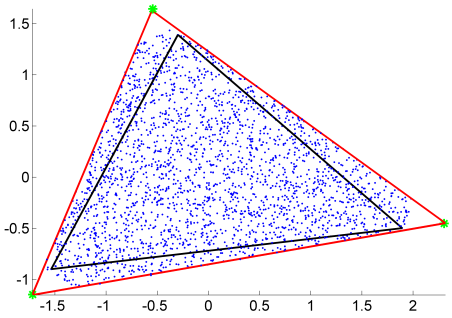
- Two unmixing algorithms have been considered for the LMM. The first strategy extracts the endmembers from the whole image using the N-FINDR algorithm (Winter, 1999) and estimates the abundances using the FCLS algorithm (Heinz and C.-I Chang, 2001) (it is referred to as "SLMM" for supervised LMM). The second strategy is a Bayesian algorithm which jointly estimates the endmembers and the abundance matrix (Dobigeon et al., 2009a) (it is referred to as "ULMM" for unsupervised LMM).

- Two approaches have been considered for the PPNMM. The first strategy uses the nonlinear EEA studied by Heylen et al. (2011) and the gradient-based approach based on the PPNMM studied in Section 1.4 for estimating the abundances and the nonlinearity parameter. This strategy is referred to as "SPPNMM" (supervised PPNMM). The second strategy is the proposed unmixing procedure referred to as "UPPNMM" (unsupervised PPNMM).

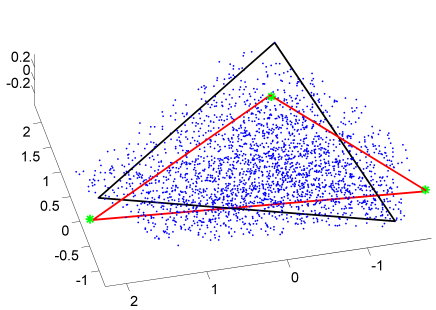
- The unmixing strategy used for the GBM is the nonlinear EEA studied in (Heylen et al., 2011) and the gradient-based algorithm presented in (Halimi et al., 2011b) for abundance estimation.

Table 1.6 shows the RNMSEs associated with the images $I_1, \ldots, I_3$ for the different estimation procedures. These results show that the proposed UPPNMM performs better (in term of RNMSE) than the other considered unmixing

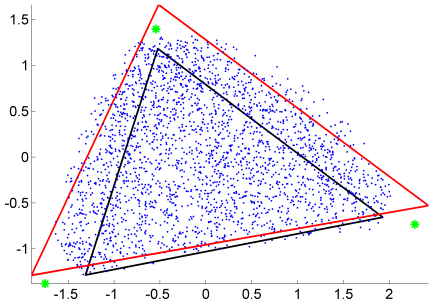
methods for the three images. Moreover, the proposed method provides similar results when compared with the ULMM for the linearly mixed image I_1.



(a) I_1



(b) I_2



(c) I_3

Figure 1.10: Visualization of the $N = 2500$ pixels (blue dots) of I_1, I_2 and I_3 using the first principal components provided by the standard PCA. The green stars correspond to the actual endmembers and the triangles are the simplexes defined by the endmembers estimated by the Heylen's method (black) and the proposed method (red).

Fig. 1.10 compares the endmember simplexes estimated by Heylen's method (Heylen et al., 2011) (black) (used to build the endmember prior) and by the proposed method (red) to the actual endmembers (green stars). For visualization, the observed pixels and the actual and estimated endmembers have been projected onto the three first axes provided by the principal component analysis. These figures show that the proposed unmixing procedure provides accurate estimated endmembers for the three images I_1 to I_3. Due to the absence of pure pixels in the image, the manifold generated by the observed pixels $\mathbf{Y}$ is difficult to estimate. This explains the limited performance obtained with Heylen's method. Conversely, the use of the prior (1.45) allows the endmembers $\mathbf{m}_r$ to depart from the prior estimations $\bar{\mathbf{m}}_r$ leading to improved performance.

The quality of endmember estimation is also evaluated by the spectral angle mapper (SAM) defined as

$$\text{SAM} = \arccos\left(\frac{\langle \hat{\mathbf{m}}_r, \mathbf{m}_r \rangle}{\|\hat{\mathbf{m}}_r\| \, \|\mathbf{m}_r\|}\right) \tag{1.71}$$

where $\mathbf{m}_r$ is the rth actual endmember and $\hat{\mathbf{m}}_r$ its estimate. The smaller |SAM|, the closer the estimated endmembers to their actual values. Table 1.7 compares the performance of the different endmember estimation algorithms. This table shows that the proposed UPPNMM generally provides more accurate endmember estimates than the other methods. Moreover, these results illustrate the robustness of the PPNMM regarding model mis-specification. Note that the ULMM and the UPPNMM provide similar results (in term of SAMs) for the image I_1 generated according to the LMM.

Table 1.7: SAMs ($\times 10^{-2}$): synthetic images.

		N-Findr	ULMM	Heylen	UPPNMM
I_1	$\mathbf{m}_1$	5.68	0.95	6.42	**0.27**
	$\mathbf{m}_2$	5.85	**0.32**	7.46	0.36
	$\mathbf{m}_3$	3.31	0.30	5.26	**0.27**
I_2	$\mathbf{m}_1$	9.27	9.68	6.71	**0.59**
	$\mathbf{m}_2$	8.58	8.67	11.80	**0.38**
	$\mathbf{m}_3$	4.47	6.34	4.98	**0.26**
I_3	$\mathbf{m}_1$	7.35	3.42	6.48	**1.50**
	$\mathbf{m}_2$	10.68	**3.13**	11.88	3.22
	$\mathbf{m}_3$	4.34	7.44	3.20	**0.85**

Table 1.8 compares the AREs obtained for the different synthetic images. These results show that the AREs are close for the different unmixing algorithms even if the estimated abundances can vary more significantly (see Table 1.6). Again, the proposed PPNMM seems to be more robust than the other mixing models to deviations from the actual model in term of ARE.

Table 1.8: AREs ($\times 10^{-2}$): synthetic images.

		I_1 (LMM)	I_2 (PPNMM)	I_3 (GBM)
LMM	SLMM	1.04	1.74	15.16
	ULMM	**0.99**	1.43	1.07
PPNMM	SPPNMM	1.26	1.27	1.31
	UPPNMM	**0.99**	**0.99**	**0.99**
GBM		1.27	1.64	1.33

As mentioned above, one of the major properties of the PPNMM is its ability to characterize the linearity/nonlinearity of the underlying mixing model for each pixel of the image via the nonlinearity parameter b_n. Fig. 1.11 shows the nonlinearity parameter distribution estimated for the three images I_1 to I_3 using the UPPNMM. This figure shows that the UPPNMM clearly identifies the linear mixtures of the image I_1 whereas more nonlinearly mixed pixels can be identified in the images I_2 and I_3. The analysis of Fig. 1.11 also shows that the nonlinearities contained in the image I_3 (GBM) are generally less significant than the nonlinearities affecting I_2 (PPNMM) for a same signal-to-noise ratio (SNR $\simeq$ 21dB).

The next set of simulations analyzes the performance of the proposed UPPNMM algorithm for different numbers

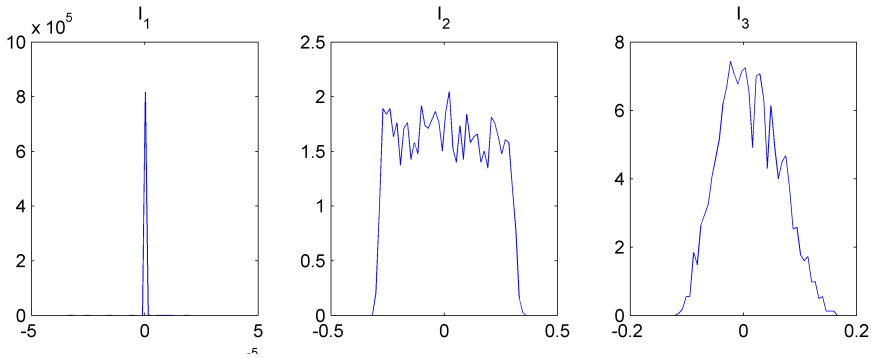


Figure 1.11: Distributions of the nonlinearity parameters b_n for the images I_1 (left), I_2 (middle) and I_3 (right).

of endmembers ($R \in \{4, 5, 6\}$) by unmixing three synthetic images of $N = 2500$ pixels distributed according to the PPNMM. The endmembers contained in these images have been extracted from the spectral libraries provided with

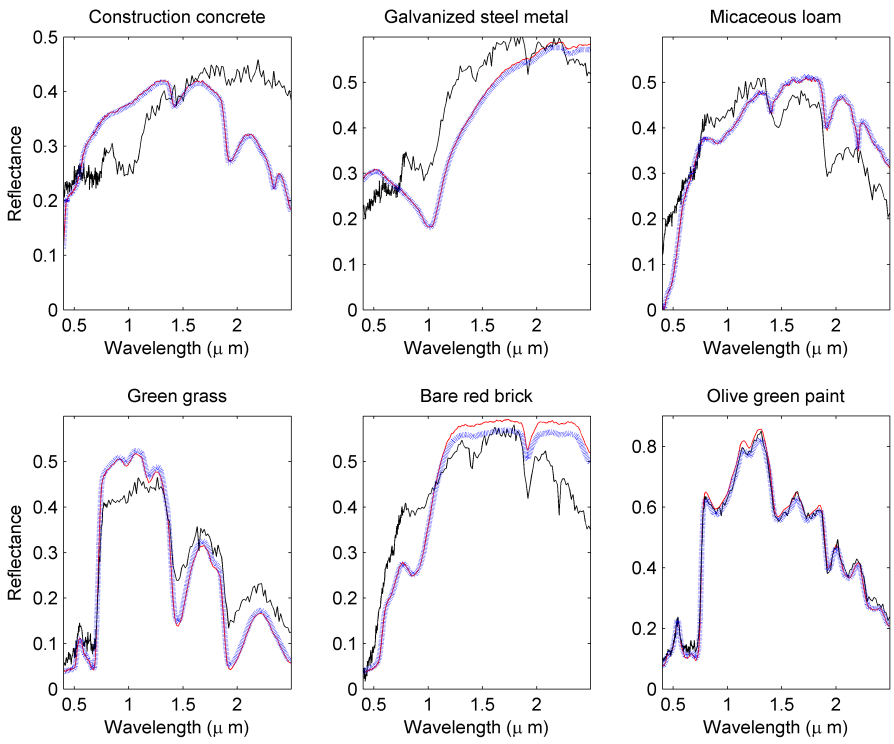


Figure 1.12: Actual endmembers (blue dots) and the endmembers estimated by Heylen's method (black lines) and the UPPNMM (red lines) for the synthetic image containing $R = 6$ endmembers.

the ENVI software (RSI (Research Systems Inc.), 2003). For each image, the abundance vectors $\boldsymbol{a}_n, n = 1, \ldots, N$ have been randomly generated according to a uniform distribution over the admissible set (1.70). All images have been corrupted by an additive white Gaussian noise corresponding to $\sigma^2 = 10^{-4}$. The nonlinearity coefficients b_n are uniformly drawn in the set $[-0.3, 0.3]$. Tables 1.9 compares the performance of the proposed method in term

of endmember estimation (average SAMs of the R endmembers), abundance estimation and reconstruction error. These results show a general degradation of the abundance and endmember estimations when R is increasing (this is intuitive since the estimator variances usually increase with the number of parameters to be estimated). However, this degradation is reasonable when compared to the Heylen's method coupled with the PPNMM-based inversion algorithm introduced in Section 1.4. The proposed algorithm still provides accurate estimates, as illustrated in Fig. 1.12 which compares the actual and estimated endmembers associated with the image containing $R = 6$ endmembers.

Table 1.9: Unmixing performance:synthetic images.

		$R = 4$	$R = 5$	$R = 6$
Average SAMs ($\times 10^{-2}$)	SPPNMM	7.76	10.78	18.53
	UPPNMM	**0.47**	**0.81**	**1.09**
RNMSEs ($\times 10^{-2}$)	SPPNMM	7.58	10.95	16.52
	UPPNMM	**0.78**	**1.23**	**1.47**
AREs ($\times 10^{-2}$)	SPPNMM	1.36	1.46	1.64
	UPPNMM	**0.99**	**0.99**	**0.99**

Simulations on real data

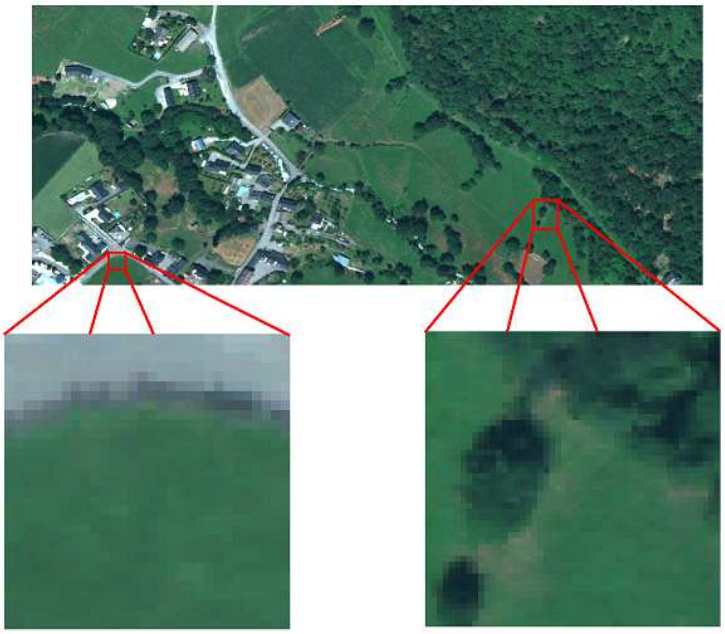


Figure 1.13: Top: real hyperspectral Madonna data acquired by the Hyspex hyperspectral scanner over Villelongue, France. Bottom: Scene #1 (left) and Scene #2 (right) shown in true colors.

The real image considered in this section was acquired in 2010 by the Hyspex hyperspectral scanner over Villelongue, France (00°03'W and 42°57'N). $L = 160$ spectral bands were recorded from the visible to near infrared with a spatial resolution of 0.5m. This dataset has already been studied in (Sheeren et al., 2011) and is mainly composed of forested

and urban areas. More details about the data acquisition and pre-processing steps are available in (Sheeren et al., 2011). Two sub-images denoted as scene #1 and scene #2 (of size 31×30 and 50×50 pixels) are chosen here to evaluate the proposed unmixing procedure and are depicted in Fig. 1.13 (bottom images). The scene #1 is mainly composed of road, ditch and grass pixels. The scene #2 is more complex since it includes shadowed pixels. For this image, shadow is considered as an additional endmember, resulting in $R = 4$ endmembers, i.e., tree, grass, soil and shadow.

The endmembers extracted by N-FINDR, the ULMM algorithm (Dobigeon et al., 2009a) and Heylen's method (Heylen et al., 2011) with $R = 3$ (resp. $R = 4$) for the scene #1 (resp. scene #2) are compared with the endmembers estimated by the UPPNMM in Fig. 1.14 (resp. Fig. 1.15). For the scene #1, the four algorithms provide similar endmember estimates whereas the estimated shadow spectra are different for the scene #2. The N-FINDR algorithm and Heylen's method estimate endmembers as the purest pixels of the observed image, which can be problematic when there is no pure pixel in the image (as it occurs with shadowed pixels in the scene #2). Conversely, the ULMM and UPPNMM methods, which jointly estimate the endmembers and the abundances seem to provide more relevant shadow spectra (of lower amplitude).

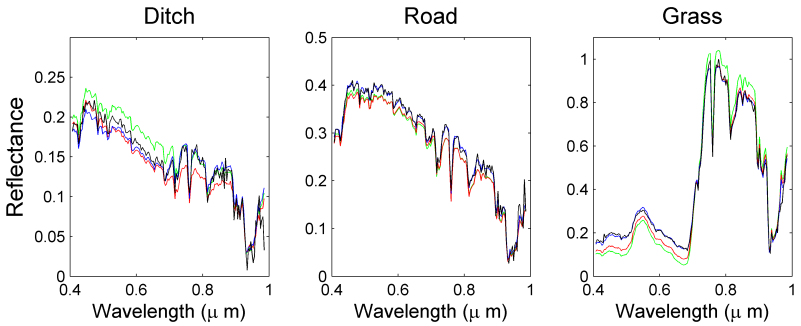


Figure 1.14: The $R = 3$ endmembers estimated by N-Findr (blue lines), ULMM (green lines), Heylen's method (black lines) and the UPPNMM (red lines) for the scene #1.

Examples of abundance maps for the scene #1 (resp. scene #2), estimated by the ULMM and the UPPNMM algorithms are presented in Fig. 1.16 (resp. Fig. 1.17). The abundance maps obtained by the UPPNMM are similar to the abundance maps obtained with ULMM.

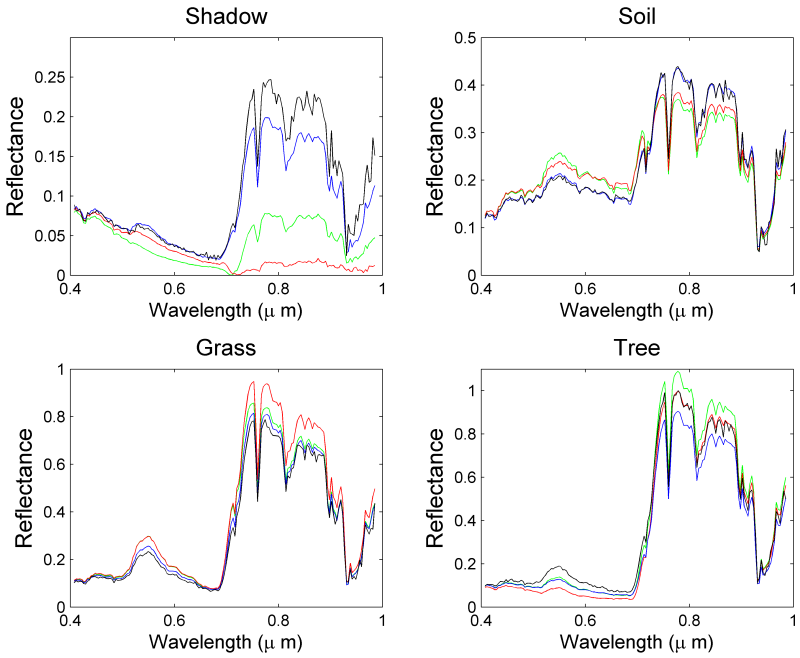


Figure 1.15: The $R = 4$ endmembers estimated by N-Findr (blue lines), ULMM (green lines), Heylen's method (black lines) and the UPPNMM (red lines) for the scene #2.

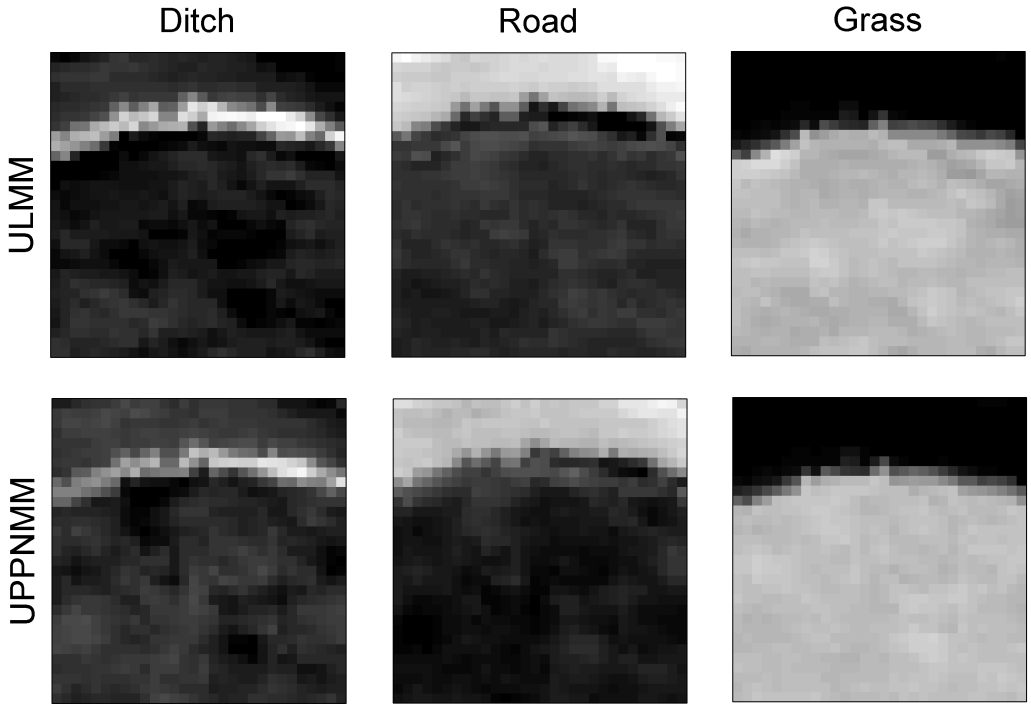


Figure 1.16: Abundance maps estimated by the SLMM, the GBM and the UPPNMM algorithms for the scene #1.

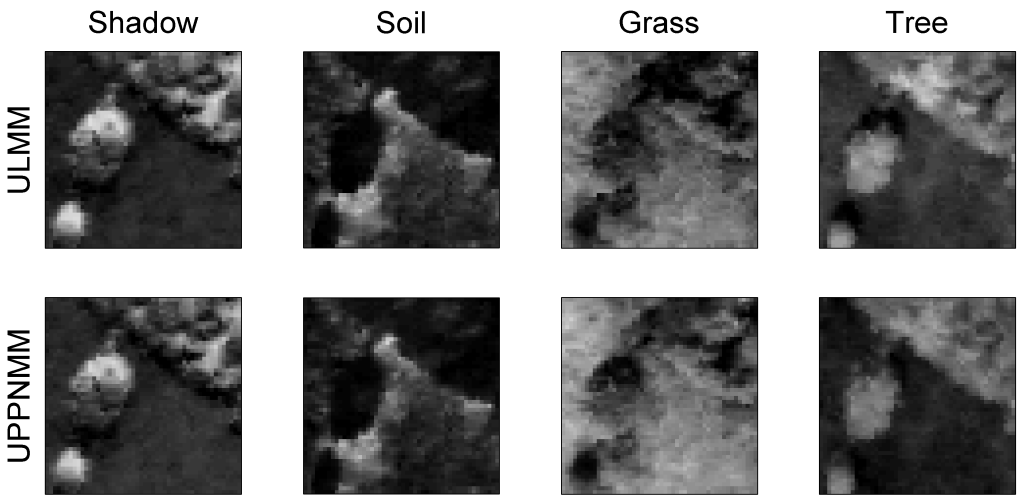
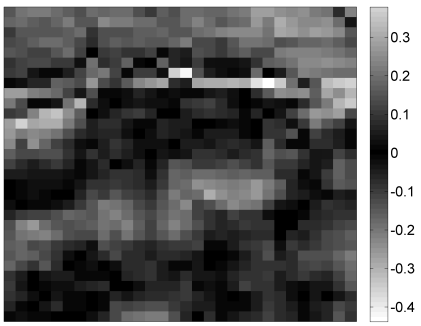
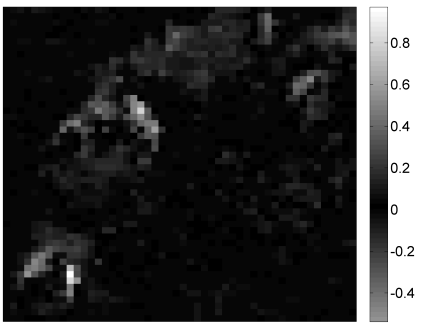


Figure 1.17: Abundance maps estimated by the SLMM, the GBM and the UPPNMM algorithms for the scene #2.

Fig. 1.18 shows the estimated maps of b_n for the two considered images. Different nonlinear regions can be identified in the scene #1, mainly in the grass-planted region (probably due to endmember variability) and near the ditch (presence of relief). For the scene #2, nonlinear effects are mainly detected in shadowed pixels.
Fig. 1.19 compares the noise variances estimated by the UPPNMM for the two real images with the noise variances estimated by the HySime algorithm (Bioucas-Dias and Nascimento, 2008). The HySime algorithm assumes additive noise and estimates the noise covariance matrix of the image using multiple regression. Fig. 1.19 first shows that the two algorithms provides similar noise variance estimates. Moreover, these results motivate the consideration of non i.i.d. noise for hyperspectral image analysis since the noise variances increase for the higher wavelengths for the two images.



(a) Scene #1



(b) Scene #2

Figure 1.18: Maps of the nonlinearity parameter b_n estimated by the UPPNMM for the real images.

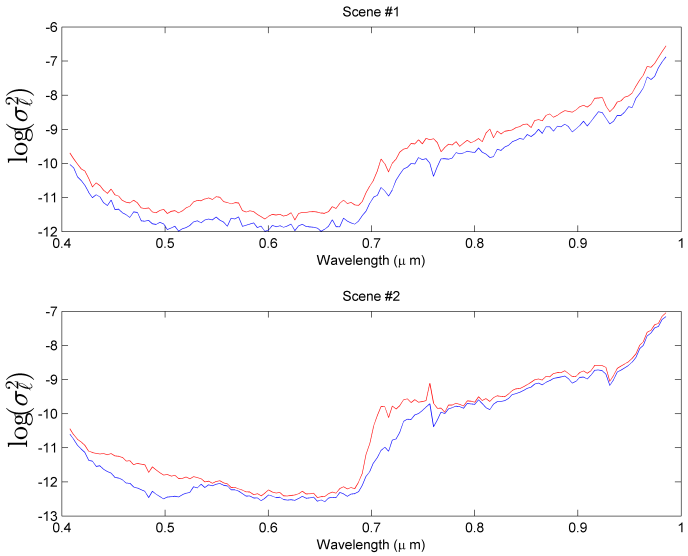


Figure 1.19: Noise variances estimated by the UPPNMM (red) and the Hysime algorithm (blue) for the scene #1 (top) and the scene #2 (bottom).

The performance of the proposed algorithm is finally evaluated by computing the AREs associated with the two real images. These AREs are compared in Table 4.6 with those obtained by assuming other mixing models. The two unsupervised algorithms (ULMM and UPPNMM) provide smaller AREs than the SU procedures decomposed into two steps. This observation motivates the use of joint abundance and endmember estimation algorithms.

Table 1.10: AREs ($\times 10^{-2}$): Real image.

		Scene #1	Scene #2
LMM	SLMM	1.53	1.04
	ULMM	1.11	**0.88**
PPNMM	SPPNMM	1.50	1.17
	UPPNMM	**1.08**	0.89
GBM		1.72	1.25

1.5.3 Intermediate conclusion

We proposed a new hierarchical Bayesian algorithm for unsupervised nonlinear spectral unmixing of hyperspectral images. This algorithm assumed that each pixel of the image is a post-nonlinear mixture of the endmembers contaminated by additive Gaussian noise. The physical constraints for the abundances and endmembers were included in the Bayesian framework through appropriate prior distributions. Due to the complexity of the resulting joint posterior distribution, a Markov chain Monte Carlo method was used to approximate the MMSE estimator of the unknown model parameters. Because of the large number of parameters to be estimated, Hamiltonian Monte Carlo methods were used to reduce the sampling procedure complexity and to improve the mixing properties of the proposed sampler. Simulations conducted on synthetic data illustrated the performance of the proposed algorithm for linear and nonlinear spectral unmixing. Finally, we would like to mention an important advantage of the proposed algorithm, i.e., its flexibility regarding the absence of pure pixels in the image.

1.6 Conclusion

In this chapter, we derived a polynomial post-nonlinear mixing model for nonlinear SU of hyperspectral images. Two scenarios have been presented, depending on the prior knowledge about the endmembers. When the endmembers are known, we proposed three unmixing algorithms for the inversion step. The proposed Bayesian model coupled with MCMC methods allowed samples to be generated according to the posterior of interest. These samples were then used to compute the classical Bayesian estimators, while providing measures of uncertainty about the parameter estimates. To overcome the computational complexity induced by the sampling procedure, two LS methods were also proposed, providing unmixing performance similar to those obtained with the Bayesian algorithm.

Estimating the endmembers associated with nonlinear mixtures is a challenging issue. In this chapter, we finally proposed a fully unsupervised unmixing strategy based on the proposed PPNMM model. This Bayesian algorithm estimated jointly the endmembers and the mixing coefficients even when there is no pure pixel in the observed image. The joint endmember and abundance estimation increases the number of parameters to be estimated which can be problematic when using standard MCMC methods such as the Gibbs sampler. To improve mixing properties of the Markov chains, efficient constrained Hamiltonian Monte Carlo methods were investigated. Simulations conducted on real data showed the accuracy of the proposed model to quantify possible nonlinear effects in each image pixel using a single parameter.

It was shown that the proposed model reduces to the classical LMM when the nonlinearity parameter equals zero. Consequently, it makes sense to use this nonlinear model for nonlinearity detection. Identifying nonlinearly from linearly mixed pixel is an interesting issue for a deeper scene understanding. This problem, in particular using the PPNMM, will be addressed in the last chapter of this manuscript.

The nonlinear SU algorithms presented in this chapter were based on a parametric model involving explicitly the abundances and the endmember matrix. This model assumes that the observations are post-nonlinear mixtures of endmembers and that the nonlinear transformation can be approximated using a second order polynomial function. For applications where the observed pixels cannot be modeled by post-nonlinear mixtures, other nonlinear models should be used. Nonparametric models have the interesting property of being highly flexible to model different nonlinearities and have already provided interesting results for SU (Broadwater and Banerjee, 2009; Broadwater et al., 2007; Chen et al., 2013b). The next chapter studies a new kernel-based model for unsupervised nonlinear SU. This model approximates the nonlinear relation between the abundances and the observations without explicit dependency on the unknown endmembers, leading to a more flexible mixing model.

Main contributions. A new nonlinear model based on post-nonlinear mixtures was proposed for linear/nonlinear SU. The nonlinearity in each pixel was characterized by a single amplitude parameter which will allow simple nonlinearity detectors to be derived in the third chapter. SU methods were proposed to address the problems of supervised and unsupervised unmixing. Efficient sampling algorithms based on Hamiltonian MCMC methods were also used to improve the mixing properties of the sampler when the endmembers are unknown (and have to be estimated).

1.7 Conclusion (in French)

Dans ce chapitre, nous avons étudié un premier modèle de mélange post-non-linéaire polynomial pour le démélange d'images hyperspectrales. Deux scénarios ont été présentés en fonction de la connaissance *a priori* sur les signatures des composants purs de l'image. Lorsque ces signatures sont connues, nous avons proposé trois algorithmes de démélange pour l'étape d'inversion (l'estimation des abondances et des paramètres des non-linéarités). Le modèle bayésien proposé, couplé avec des méthodes MCMC, permet de générer des échantillons suivant la loi a posteriori d'intérêt. Ces échantillons sont ensuite utilisés pour calculer les estimateurs bayésiens classiques, tout en offrant des mesures d'incertitude sur les estimations des paramètres. Pour pallier le coût calculatoire induit par la procédure d'échantillonnage, deux méthodes d'optimisation sous contraintes ont également été proposées, offrant des performances de démélange similaires à celles obtenues avec l'algorithme bayésien.

L'estimation des signatures spectrales des composants purs associées à des mélanges non-linéaires est une question difficile. Dans ce chapitre, nous avons proposé une stratégie de démélange non-supervisée basée sur le modèle PPNMM. Cet algorithme bayésien estime conjointement les spectres des composants purs et les coefficients de mélange, même lorsqu'il n'y a pas de pixel pur dans l'image observée. Une telle estimation augmente le nombre de paramètres à générer ce qui peut être problématique lors de l'utilisation des méthodes MCMC standards tels que l'échantillonneur de Gibbs. Pour améliorer les propriétés de mélange des chaînes de Markov, des méthodes de Monte Carlo à dynamiques hamiltoniennes contraintes efficaces ont été étudiées. Les simulations effectuées sur données réelles ont montré la pertinence du modèle proposé pour quantifier les effets non-linéaires possibles dans chaque pixel de l'image à l'aide d'un seul paramètre.

Il a été montré que le modèle proposé se réduit au modèle de mélange linéaire classique lorsque le paramètre de non-linéarité est égal à zéro. Par conséquent, il paraît logique d'utiliser ce modèle non-linéaire pour la détection de non-linéarités, sujet qui sera abordé dans les derniers chapitres de ce manuscrit. En effet, l'identification de pixels résultant de mélanges linéaires et non-linéaires est une question intéressante pour une caractérisation plus fine de l'image observée.

Les algorithmes de démélange non-linéaires présentés dans ce chapitre sont fondés sur un modèle paramétrique impliquant explicitement les abondances et la matrice des signatures spectrales des composants de l'image. Ce modèle suppose que les observations sont des mélanges post-non-linéaires des spectres des composants de l'image et que la transformation non-linéaire peut être estimée à l'aide d'un polynôme du second ordre. Pour les applications où les pixels observés ne peuvent pas être modélisés par des mélanges post-non-linéaires, d'autres modèles non-linéaires devraient être utilisés. Les modèles non-paramétriques ont la propriété intéressante d'être très flexible pour modéliser différentes non-linéarités et ont déjà donné des résultats intéressants pour le démélange spectral (Broadwater and Banerjee, 2009; Broadwater et al., 2007; Chen et al., 2013b). Le chapitre suivant étudie un nouveau modèle non-linéaire à base de noyaux pour le démélange non-linéaire non-supervisé. Ce modèle approche la relation non-linéaire entre les abondances et les observations sans dépendre explicitement des spectres des composants purs de l'image, conduisant à un modèle de mélange plus souple.

Contributions majeures. Un nouveau modèle non-linéaire à base de mélanges post-non-linéaires a été proposé pour le démélange spectral linéaire et non-linéaire. Les non-linéarités dans chaque pixel sont caractérisées par un paramètre d'amplitude unique, ce qui permet de proposer un détecteur de non-linéarités simple qui sera étudié dans le troisième chapitre de ce manuscrit. Des méthodes de démélange spectral ont été proposées pour résoudre les problèmes de démélange supervisé et non-supervisé. Des algorithmes d'échantillonnage efficaces basés sur des

méthodes de Monte Carlo à dynamiques hamiltoniennes ont également été utilisés pour améliorer les propriétés de mélange de l'échantillonneur lorsque les signatures spectrales des composants purs ne sont pas connues (et doivent être estimées).

Chapter 2

Unsupervised nonlinear unmixing using Gaussian processes

This chapter has been adapted from the journal paper Altmann et al. (2013a) (published).

Contents

2.1 Introduction (in French)

Dans ce chapitre, nous proposons une nouvelle méthode à noyau pour le démélange non-linéaire non-supervisée basée sur un modèle à variables latentes. Dans le chapitre 1, nous avons proposé un modèle post-non-linéaire pour approcher la fonction non-linéaire reliant les signatures spectrales des composants purs et leurs abondances associées

aux pixels observés. Il a également été montré que le problème de démélange non-supervisé est un problème difficile en raison du grand nombre de paramètres inconnus à estimer. Comme il est présenté dans ce chapitre, un des principaux avantages du modèle de variable latente proposé est qu'il ne repose pas explicitement sur la matrice des signatures spectrales, mais sur un nombre réduit de paramètres à estimer. De plus, la relation non-linéaire entre les abondances et les observations est estimée à l'aide d'un noyau, contrairement au modèle PPNMM. Les méthodes à noyaux ont été intensivement étudiées dans la communauté de l'apprentissage automatique pour leur capacité à modéliser des non-linéarités complexes. Broadwater et al. (2007) et Broadwater and Banerjee (2009) remplacent les produits scalaires utilisés pour le démélange linéaire par des noyaux afin de généraliser l'algorithme classique FCLS (Heinz and C.-I Chang, 2001) en Kernel-FCLS (K-FCLS). L'algorithme K-FCLS suppose que les spectres des composants purs sont connus et résout le problème de démélange non-linéaire en projetant les données dans un espace de grande dimension (généralement supérieure à la dimension des données) et en résolvant le problème de démélange linéaire dans ce nouvel espace. L'algorithme K-FCLS appartient à la même famille de méthode que l'analyse en composantes principales non-linéaire (K-PCA) dans le sens où ces deux méthodes sont basées sur des modèles discriminants. Physiquement, modeliser les effets non-linéaires qui rentrent en jeu dans les processus de mélanges par un modèle discriminant n'est pas intuitif. Au contraire, il parait plus naturel de considérer des modèles génératifs qui relient la composition physique de chaque pixel à leur spectre associé. De tels modèles peuvent s'écrire de la façon suivante

$$\mathbf{y}_n \approx \varphi_{\mathbf{M}}(\boldsymbol{a}_n, \boldsymbol{\theta}_n)$$

où $\varphi_{\mathbf{M}}(\cdot)$ est une transformation (non-linéaire) paramétrée par la matrice $\mathbf{M}$ des signatures spectrales des composants de l'image, $\boldsymbol{a}_n$ est le vecteur d'abondance du $n$ième pixel et $\boldsymbol{\theta}_n$ est un vecteur qui contient de possibles paramètres supplémentaires associés au $n$ième pixel. Il est important de noter que les modèles génératifs présentés dans l'introduction de ce manuscrit (comme les modèles bilinéaires) peuvent s'écrire sous cette forme.

Pour de nombreuses images réelles, on observe que les données vivent souvent sur une variété dont la dimension est étroitement liée au nombre de composants présents dans l'image. Plus précisement, due à la contrainte d'égalité (somme-à-un) des abondances, la dimension de cette variété est souvent $(R-1)$, où R est le nombre de composants de l'image. Dans le cas de mélanges linéaires, cette variété est incluse dans un sous-espace de dimension $(R-1)$, ce qui peut être utilisé pour identifier le nombre de composants de la scène. Dans le cas de mélanges non-linéaires, cette variété est incluse dans un sous-espace de dimension supérieure. Dans ce chapitre, on suppose donc que les données (dans le cas sans bruit) vivent sur une variété (caractérisée par les composants de l'image) et que la position des observations sur cette variété dépend seulement des abondances de ces matériaux.

L'algorithme de démélange proposé se décompose en trois étapes. La première étape, dite de réduction de dimension, consiste à identifier la variété sur laquelle vivent les données et à estimer les coordonnées des pixels observés sur cette variété par l'intermédiaire de variables latentes. La deuxième étape permet d'identifier les abondances associées à chaque pixel et reliées aux variables latentes. Enfin, la dernière étape consiste à prédire les signatures spectrales des composants de l'image à l'aide de processus gaussiens. L'algorithme de démélange non-supervisé proposé est comparé à des algorithmes de démélange linéaires et non-linéaires de la littérature par des simulations sur données synthétiques et réelles. Ces simulations permettent également de mettre en évidence la flexibilité du modèle considéré pour résoudre le problème de démélange non-linéaire quand seul le nombre de composants est connu, même s'il n'y a pas de pixel pur dans l'image.

2.2 Introduction

In this chapter we propose a new kernel-based method for unsupervised nonlinear SU based on a latent variable model. In Chapter 1, we proposed a post-nonlinear model to approximate the nonlinear function relating the endmembers and the abundances to the observed pixels. It has been shown that the unsupervised unmixing problem is difficult because of the large number of unknown parameters to be estimated. As it will be shown later in this chapter, one of the main advantages of the proposed latent variable model is that it does not rely explicitly on the endmember matrix, but on a reduced number of unknown parameters instead. Moreover, conversely to the PPNMM, the nonlinear relation between the abundances and the observations is approximated using a kernel function. Kernel-based methods have been intensively studied in the machine learning community for their ability to model complex nonlinearities. Broadwater and Banerjee (2009); Broadwater et al. (2007) propose to replace inner products used for linear SU by kernel functions to generalize the classical FCLS algorithm (Heinz and C.-I Chang, 2001), yielding the Kernel-FCLS (K-FCLS) algorithm. The K-FCLS algorithm assumes that the endmembers are known and solves the following constrained problem

$$\min_{\boldsymbol{a}_n} \left[\left(\phi(\mathbf{y}_n) - \sum_{r=1}^{R} a_{r,n} \phi(\mathbf{m}_r) \right)^T \left(\phi(\mathbf{y}_n) - \sum_{r=1}^{R} a_{r,n} \phi(\mathbf{m}_r) \right) \right] \tag{2.1}$$

subject to the positivity and sum-to-one constraints for the abundance vectors $\boldsymbol{a}_n$, for $n = 1, \ldots, N$. The nonlinear transformation $\phi(\cdot)$ is applied to the L-dimensional observations $\mathbf{y}_n$ such that the set $(\phi(\mathbf{y}_n))_{n=1,\ldots,N}$ lies into a subspace whose dimension is (much) larger than L (potentially of infinite dimension). Solving (2.1) consists of projecting the transformed observations onto the simplex whose vertices are the transformed endmembers $\phi(\mathbf{m}_1), \ldots, \phi(\mathbf{m}_R)$.

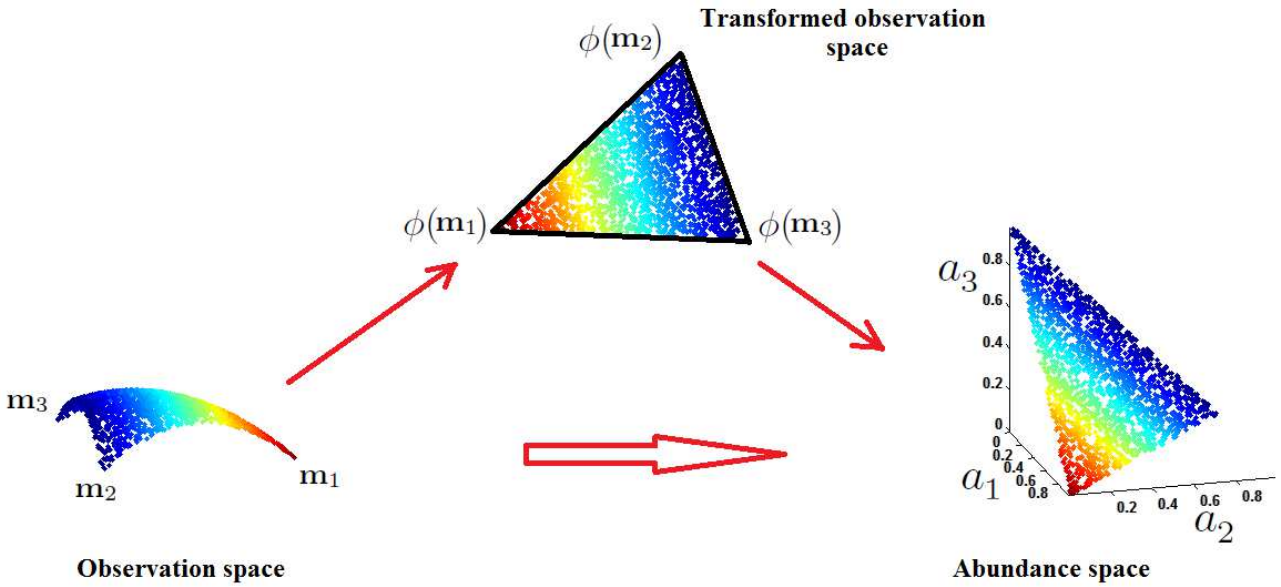


Figure 2.1: K-FCLS algorithm principle. The K-FCLS approximates the nonlinear mapping relating the observations to the abundances.

Fig. 2.1 illustrates the K-FCLS method principle for $R = 3$ endmembers. The K-FCLS algorithm can be interpreted

as a nonlinear dimensionality reduction technique, similar to the Kernel principal component analysis (KPCA), in the sense that a nonlinear transformation in applied to the observations to recover the abundances using

$$\boldsymbol{a}_n \approx \varphi_{\mathbf{M}}(\mathbf{y}_n)$$

where $\varphi_{\mathbf{M}}(\cdot)$ is a nonlinear transformation parameterized by the endmember matrix $\mathbf{M}$ that is approximated using kernel functions. Physically, modeling nonlinearities by setting a nonlinear mapping from the observations to the abundances is not intuitive. Conversely, it seems more natural to assume a nonlinear mapping from the pixel compositions (i.e., the abundances and endmembers) to the observations. Such mappings can be expressed as

$$\mathbf{y}_n \approx \varphi_{\mathbf{M}}(\boldsymbol{a}_n, \boldsymbol{\theta}_n) \tag{2.2}$$

where $\boldsymbol{\theta}_n$ contains possible additional parameters associated with the nth pixel. It is important to note that the generative models presented in the introduction of this manuscript (such as the bilinear models) can be expressed as (2.2) where $\boldsymbol{\theta}_n$ contains nonlinearity parameters.

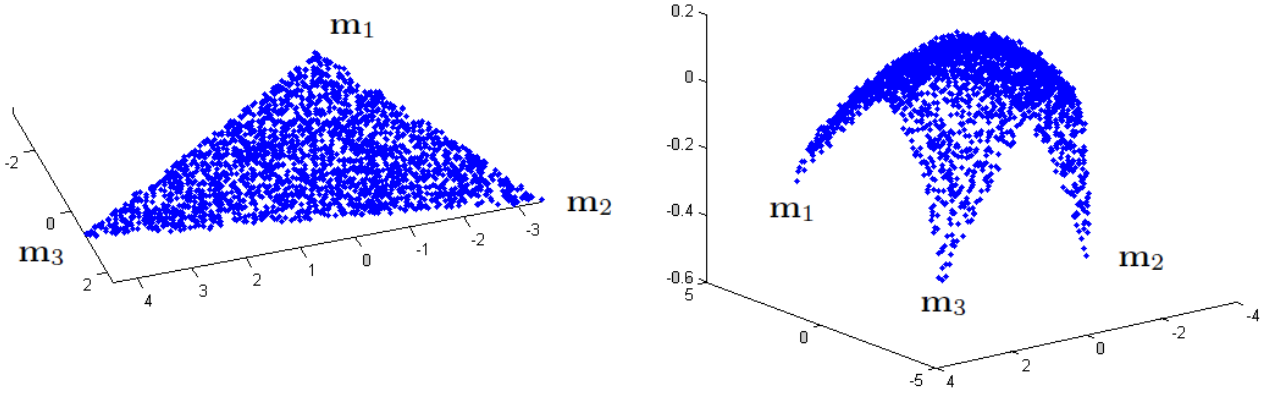


Figure 2.2: Representation of mixtures composed of $R = 3$ components distributed according to the LMM (left) and the FM (right), using the first three principal components provided by the standard PCA.

For linearly mixed pixels composed of R materials (in the noise-free case), the observed pixels belong to an $(R-1)$-dimensional subspace that can be identified using principal subspace methods (such as principal component analysis (PCA)) (see Fig. 2.2 (left)). Conversely, when nonlinear mixtures occur, the dimension of the principal subspace containing the data increases with the nonlinearity complexity. In this chapter, we assume however that mixed pixels still belong to an $(R-1)$-dimensional manifold when R endmembers are involved in the mixtures. This manifold is characterized by the mixture compositions (abundances) and can be either linear (for linearly mixed pixels) or nonlinear and included within a higher dimensional subspace. This is typically the case for intimate mixtures models and the bilinear FM (Fan et al., 2009) (see Fig. 2.2 (right)). Based on this assumption, the mapping (2.2) reduces to

$$\mathbf{y}_n \approx \mathbf{g}(\boldsymbol{a}_n). \tag{2.3}$$

where $\mathbf{g}(\cdot)$ is a linear/nonlinear transformation that implicitly relies on the endmember matrix $\mathbf{M}$. The resulting latent variable model (LVM) approximates the observed data using a nonlinear transformation of the abundance,

contrary to KPCA-based methods which approximate the abundances using nonlinear transformations of the observations (Fig. 2.3). Conversely to the K-FCLS algorithm, this chapter addresses the problem of unsupervised unmixing, i.e., the endmembers are assumed to be unknown. Precisely, only the number of components is known. Consequently, the nonlinear mapping $\mathbf{g}(\cdot)$ and the abundances in (2.3) are unknown and have to be estimated. The next paragraph presents a new nonlinear mixing model based on LVMs for unsupervised SU.

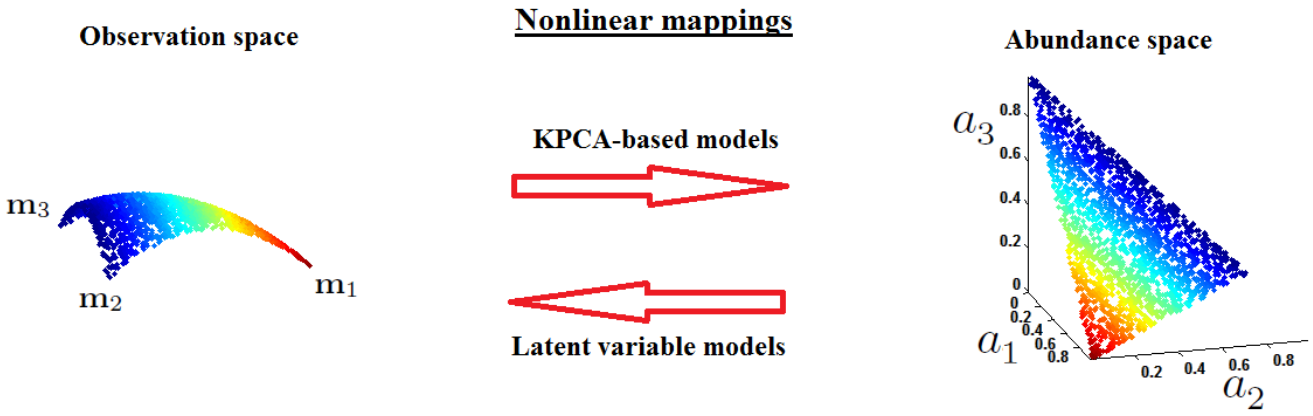


Figure 2.3: Modeling nonlinear relations between observed pixels and corresponding abundances.

2.3 Nonlinear mixing model

Consider a hyperspectral image of N pixels, composed of R endmembers observed in L spectral bands. For convenience, the data are assumed to have been previously centered, i.e., the sample mean of the N original pixels has been subtracted from each observed pixel. The L-spectrum $\mathbf{y}_n = [y_{1,n}, \ldots, y_{L,n}]^T$ of the nth mixed pixel ($n = 1, \ldots, N$) is defined as a transformation of its corresponding abundance vector $\boldsymbol{a}_n = [a_{1,n}, \ldots, a_{R,n}]^T$ as follows

$$\mathbf{y}_n = \mathbf{g}(\boldsymbol{a}_n) + \mathbf{e}_n, \quad n = 1, \ldots, N \tag{2.4}$$

where $\mathbf{g} : \mathbb{R}^R \to \mathbb{R}^L$ is a linear or nonlinear unknown function. The noise vector $\mathbf{e}_n$ is an i.i.d. white Gaussian noise sequence with variance σ^2, i.e., $\mathbf{e}_n \sim \mathcal{N}\left(\mathbf{e}_n | \mathbf{0}_L, \sigma^2 \mathbf{I}_L\right), n = 1, \ldots, N$. Without loss of generality, the nonlinear mapping (2.4) from the abundance space to the observation space can be rewritten

$$\mathbf{y}_n = \mathbf{W}_0 \boldsymbol{\psi}[\boldsymbol{a}_n] + \mathbf{e}_n, \quad n = 1, \ldots, N \tag{2.5}$$

where $\boldsymbol{\psi} : \mathbb{R}^R \to \mathbb{R}^D$, $\mathbf{W}_0$ is an $L \times D$ matrix and the dimension D is the dimension of the subspace spanned by the transformed abundance vectors $\boldsymbol{\psi}[\boldsymbol{a}_n], n = 1, \ldots, N$. Of course, the performance of the unmixing strategy relies on the choice of the nonlinear function $\boldsymbol{\psi}$. In this study, we will use the following nonlinearity

$$\begin{aligned} \boldsymbol{\psi} : \quad & \mathbb{R}^R \to \mathbb{R}^D \\ & \boldsymbol{a} \mapsto \boldsymbol{\psi}[\boldsymbol{a}] = [a_1, \ldots, a_R, a_1 a_2 \ldots, a_{R-1} a_R]^T, \end{aligned} \tag{2.6}$$

with $D = R(R+1)/2$. It is important to note from (2.5) and (2.6) that $\mathbf{W}_0$ contains the R spectra $\mathbf{m}_1, \ldots, \mathbf{m}_R$ of the pure components present in the image and $R(R-1)/2$ interaction spectra $\mathbf{t}_{R+1}, \ldots, \mathbf{t}_D$ between these components,

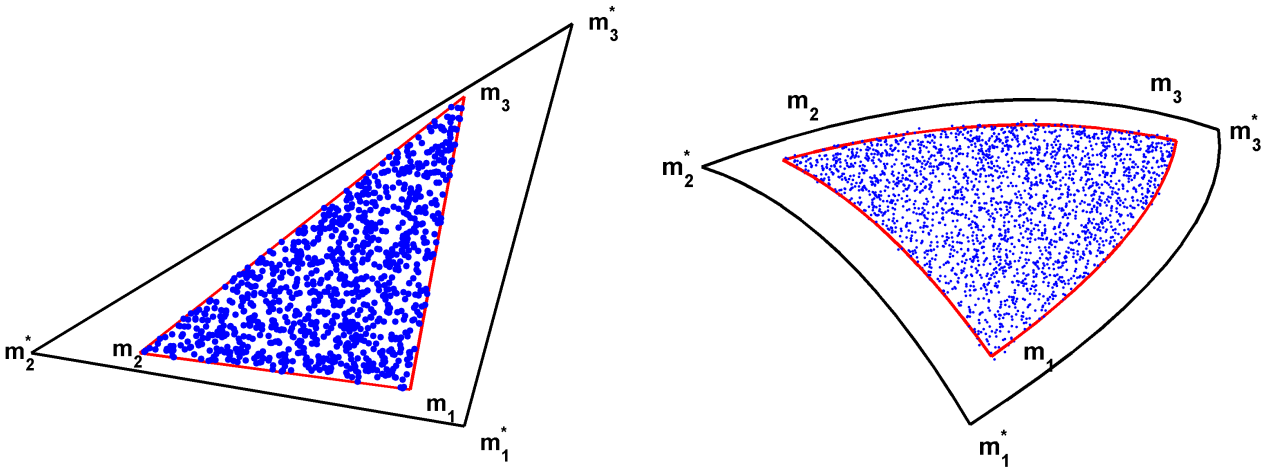
i.e., $\mathbf{W}_0 = [\mathbf{m}_1, \ldots, \mathbf{m}_R, \mathbf{t}_{R+1}, \ldots, \mathbf{t}_D]$. The primary motivation for considering this particular kind of nonlinearity is the fact that the resulting mixing model is a bilinear model with respect to each abundance $a_r, r = 1, \ldots, R$. More precisely, this mixing model reduces to the generalized bilinear model proposed in (Halimi et al., 2011a) for

$$\mathbf{t}_{i+j} = \gamma_{i,j}\mathbf{m}_i \odot \mathbf{m}_j, \quad \forall j > i, \tag{2.7}$$

and thus the LMM. Note also that the analysis in this chapter could be applied to any other nonlinearity $\boldsymbol{\psi}$.
Due to physical constraints, the abundance vector $\boldsymbol{a}_n$ satisfies the following positivity and sum-to-one constraints

$$\sum_{r=1}^{R} a_{r,n} = 1, \quad a_{r,n} \geq 0, \forall r \in \{1, \ldots, R\}. \tag{2.8}$$

Since the nonlinearity $\boldsymbol{\psi}$ is fixed, the problem of unsupervised spectral unmixing is to determine the $L \times D$ spectrum matrix $\mathbf{W}_0$, the $R \times N$ abundance matrix $\mathbf{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N]$ satisfying (2.5) with the constraints (2.8), and the noise variance σ^2.



(a) LMM (b) FM

Figure 2.4: Representation of pixels (blue dots) generated according to the LMM (a) and the FM (b) and admissible endmembers solving the unmixing problem.

Unfortunately, it is well known that the unmixing problem is ill-posed and that the solution of this constrained problem is not unique even when considering the constraints (2.8) for the abundances. In the noise-free linear case, it is well known that the data are contained in a simplex whose vertices are the endmembers. Fig. 2.4 (left) shows two admissible sets of endmembers (i.e., $\{\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3\}$ and $\{\mathbf{m}_1^*, \mathbf{m}_2^*, \mathbf{m}_3^*\}$) which can generate the linearly mixed pixels (blue dots). When estimating the endmembers in the linear case, a simplex of minimum volume embedding the data is expected (i.e., the set $\{\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3\}$ is expected). Equivalently, the estimated abundance vectors are expected to occupy the largest volume in the simplex defined by (2.8). Fig. 2.4 (right) shows a set of pixels distributed according to the FM (Fan et al., 2009) for $R = 3$ endmembers. The sets $\{\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3\}$ and $\{\mathbf{m}_1^*, \mathbf{m}_2^*, \mathbf{m}_3^*\}$ (and the two sets of associated abundances) are admissible solutions for the nonlinear SU problem. Similarly to the linear case, the estimated abundance matrix resulting from an unsupervised nonlinear SU strategy is expected to occupy the largest volume in the simplex defined by (2.8) (corresponding to the endmembers $\mathbf{m}_1, \mathbf{m}_2$ and$\mathbf{m}_3$). However, assigning an appropriate prior distribution enforcing the abundances to occupy the largest volume for nonlinear

unmixing is challenging. Note that it has been shown in (Dobigeon et al., 2009a) that for linear unmixing, assigning a uniform distribution on the simplex defined by (2.8) for the abundances favors the abundances to occupy a large volume in (2.8). However, such prior can be too uninformative for nonlinear unmixing.

To tackle this problem, we first propose to relax the positivity constraints for the elements of the matrix $\mathbf{A}$ and to consider only the sum-to-one constraint. For ease of understanding, we introduce $R \times 1$ vectors satisfying the sum-to-one constraint

$$\sum_{r=1}^{R} x_{r,n} = 1, \quad n = 1, \ldots, N \tag{2.9}$$

referred to as *latent variables* and denoted as $\boldsymbol{x}_n = [x_{1,n}, \ldots, x_{R,n}]^T, n = 1, \ldots, N$. Relaxing the positivity constraints allows the problem complexity to be reduced by 1) estimating the latent variables introduced artificially and 2) scaling subsequentially these variables to recover abundances occupying the largest volume in the domain defined by (2.8). The scaling procedure will be discussed in paragraph 2.5. The next paragraph presents the Bayesian model for latent variable estimation using GPLVMs.

2.4 Bayesian model

GPLVMs (Lawrence, 2003) are powerful tools for probabilistic nonlinear dimensionality reduction that rewrite the nonlinear model (2.4) as a nonlinear mapping from a latent space to the observation space as follows

$$\mathbf{y}_n = \mathbf{W}\boldsymbol{\psi}\left[\boldsymbol{x}_n\right] + \mathbf{e}_n, \quad n = 1, \ldots, N \tag{2.10}$$

where $\boldsymbol{\psi}$ is defined in (2.6), $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_L]^T$ is an $L \times D$ matrix with $\mathbf{w}_\ell = [w_{\ell,1}, \ldots, w_{\ell,D}]^T$, and $D = R(R+1)/2$. Note that from (2.5) and (2.10) the columns of $\mathbf{W}$ span the same subspace as the columns of $\mathbf{W}_0$. Consequently, the columns of $\mathbf{W}$ are linear combinations of the spectra of interest, i.e., the columns of $\mathbf{W}_0$. Note also that when $\mathbf{W}$ is full rank, it can be shown that the latent variables are necessarily linear combinations of the abundance vectors of interest (see Appendix D for details). Figs. 2.5 and 2.6 illustrate the mapping from the abundance vectors to the observations that will be used in this study. Note that the linear mapping between the abundances and the latent variables will be explained in details in paragraph 2.5. The $D \times 1$ vectors $\boldsymbol{\psi}\left[\boldsymbol{x}_n\right]$ will be denoted as $\boldsymbol{\psi}_x(n)$ in the sequel.

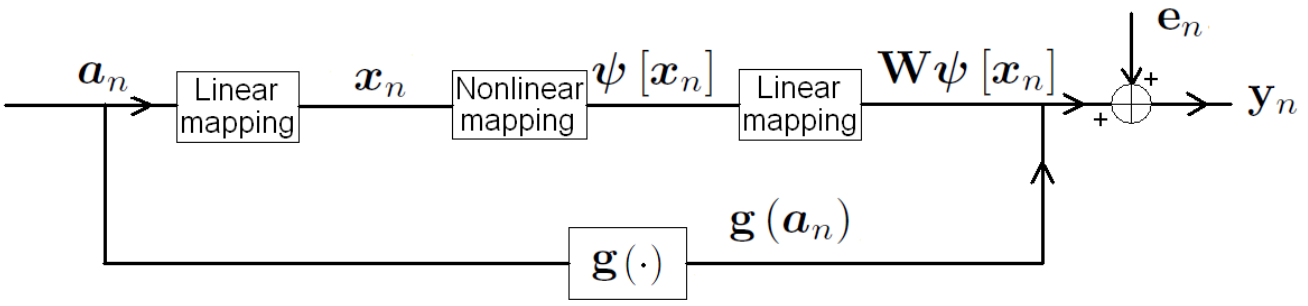


Figure 2.5: Nonlinear mapping from the abundances vectors to the observed mixed pixels.

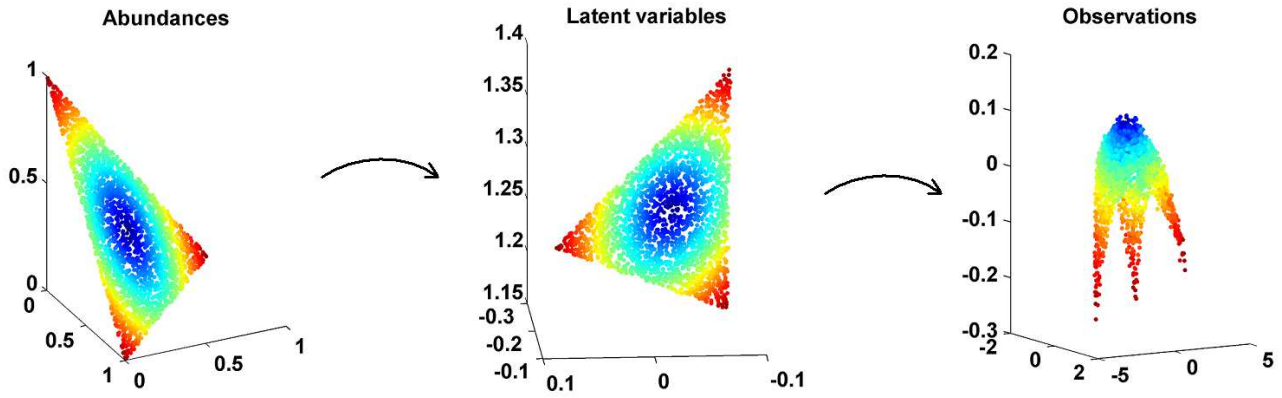


Figure 2.6: Example of mapping decomposition from the abundance vectors to the observed nonlinearly mixed pixels through the latent variables ($R = 3$).

Assuming independence between the observations, the statistical properties of the noise lead to the following likelihood of the $L \times N$ observation matrix $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N]$

$$\mathbf{Y}|\mathbf{W}, \mathbf{X}, \sigma^2 \sim \prod_{n=1}^{N} \mathcal{N}\left(\mathbf{y}_n|\mathbf{W}\boldsymbol{\psi}_x(n), \sigma^2\mathbf{I}_L\right) \tag{2.11}$$

where $\mathbf{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]$ is the $R \times N$ latent variable matrix. Note that the likelihood can be rewritten as a product of Gaussian distributions over the spectral bands as follows

$$\mathbf{Y}|\mathbf{W}, \mathbf{X}, \sigma^2 \sim \prod_{\ell=1}^{L} \mathcal{N}\left(\mathbf{y}_{\ell,:}|\boldsymbol{\Psi}_x\mathbf{w}_\ell, \sigma^2\mathbf{I}_L\right) \tag{2.12}$$

where $\mathbf{Y} = [\mathbf{y}_{1,:}, \ldots, \mathbf{y}_{L,:}]^T$ ($\mathbf{y}_{\ell,:}$ denotes the ℓth row of $\mathbf{Y}$) and $\boldsymbol{\Psi}_x = [\boldsymbol{\psi}_x(1), \ldots, \boldsymbol{\psi}_x(N)]^T$ is an $N \times D$ matrix. The idea of GPLVMs is to consider $\mathbf{W}$ as a nuisance parameter, to assign a Gaussian prior to $\mathbf{W}$ and to marginalize the joint likelihood (2.11) over $\mathbf{W}$, i.e.,

$$f(\mathbf{Y}|\mathbf{X}, \sigma^2) = \int f(\mathbf{Y}|\mathbf{W}, \mathbf{X}, \sigma^2) f(\mathbf{W}) \mathrm{d}\mathbf{W} \tag{2.13}$$

where $f(\mathbf{W})$ is the prior distribution of $\mathbf{W}$. The estimation of $\mathbf{X}$ and σ^2 can then be achieved by maximizing (2.13) following the maximum likelihood estimator (MLE) principle. An alternative consists of using an appropriate prior distribution $f(\mathbf{X}, \sigma^2)$, assuming prior independence between $\mathbf{W}$ and $(\mathbf{X}, \sigma^2)$, and maximizing the joint posterior distribution

$$\begin{aligned} f(\mathbf{X}, \sigma^2|\mathbf{Y}) &\propto \int f(\mathbf{Y}|\mathbf{W}, \mathbf{X}, \sigma^2) f(\mathbf{W}) f(\mathbf{X}, \sigma^2) \mathrm{d}\mathbf{W} \\ &\propto f(\mathbf{X}, \sigma^2) \int f(\mathbf{Y}|\mathbf{W}, \mathbf{X}, \sigma^2) f(\mathbf{W}) \mathrm{d}\mathbf{W} \\ &\propto f(\mathbf{Y}|\mathbf{X}, \sigma^2) f(\mathbf{X}, \sigma^2) \end{aligned} \tag{2.14}$$

with respect to (w.r.t.) $(\mathbf{X}, \sigma^2)$, yielding the maximum a posteriori (MAP) estimator of $(\mathbf{X}, \sigma^2)$. The next paragraph discusses different possibilities for marginalizing the joint likelihood (2.12) w.r.t. $\mathbf{W}$.

2.4.1 Marginalizing W

It can be seen from (2.13) that the marginalized likelihood and thus the associated latent variables depend on the choice of the prior $f(\mathbf{W})$. More precisely, assigning a given prior for $\mathbf{W}$ favors particular representations of the data, i.e., particular solutions for the latent variable matrix $\mathbf{X}$ maximizing the posterior (2.14). When using GPLVMs for dimensionality reduction, a classical choice (Lawrence, 2003) consists of assigning independent Gaussian priors for $\mathbf{w}_1, \ldots, \mathbf{w}_L$, leading to

$$f(\mathbf{W}) = \left(\frac{1}{2\pi}\right)^{\frac{DL}{2}} \prod_{\ell=1}^{L} \exp\left[-\frac{1}{2}\left\|\mathbf{w}_\ell\right\|^2\right]. \tag{2.15}$$

However, this choice can be inappropriate for SU. First, Eq. (2.15) can be incompatible with the admissible latent space, constrained by (2.9). Second, the prior (2.15) assumes the columns of $\mathbf{W}$ (linear combinations of the spectra of interest) are *a priori* Gaussian, which is not relevant for real spectra in most applications. A more sophisticated choice consists of considering *a priori* correlation between the columns (inter-spectra correlation) and rows (inter-bands correlation) of $\mathbf{W}$ using a structured covariance matrix to be fixed or estimated. In particular, introducing correlation between close spectral bands is of particular interest in hyperspectral imagery. Structured covariance matrices have already been considered in the GP literature for vector-valued kernels (Bonilla et al., 2007) (see (Álvarez et al., 2012) for a recent review). However, computing the resulting marginalized likelihood usually requires the estimation of the structured covariance matrix and the inversion of an $NL \times NL$ covariance matrix[1], which is prohibitive for SU of hyperspectral images since several hundreds of spectral bands are usually considered when analyzing real data. Sparse approximation techniques might be used to reduce this computational complexity (see (Quiñonero-candela et al., 2005) for a recent review). However, to our knowledge, these techniques rely on the inversion of matrices bigger than $N \times N$ matrices. The next paragraph presents an alternative that only requires the inversion of an $D \times D$ covariance matrix without any approximation.

2.4.2 Subspace identification

It can be seen from (2.10) that in the noise-free case, the data belong to a D-dimensional subspace that is spanned by the columns of $\mathbf{W}$. To reduce the computational complexity induced by the marginalization of the matrix $\mathbf{W}$ while considering correlations between spectral bands, we propose to marginalize a basis of the subspace spanned by $\mathbf{W}$ instead of $\mathbf{W}$ itself. More precisely, $\mathbf{W}$ can be decomposed as follows

$$\mathbf{W} = \mathbf{P}\mathbf{U}^T \tag{2.16}$$

where $\mathbf{P} = [\mathbf{p}_1, \ldots, \mathbf{p}_L]^T$ is an $L \times D$ matrix ($\mathbf{p}_\ell$ is $D \times 1$ vector) whose columns are arbitrary basis vectors of the D-dimensional subspace that contains the subspace spanned by the columns of $\mathbf{W}$ and $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_D]^T$ is a $D \times D$ matrix that scales the columns of $\mathbf{P}$. Note that the subspaces spanned by $\mathbf{P}$ and $\mathbf{W}$ are the same when $\mathbf{W}$ is full rank, resulting in a full rank matrix $\mathbf{U}$. The joint likelihood (2.12) can be rewritten as

$$\mathbf{Y}|\mathbf{P}, \mathbf{U}, \mathbf{X}, \sigma^2 \sim \prod_{\ell=1}^{L} \mathcal{N}\left(\mathbf{y}_{\ell,:}|\mathbf{C}\mathbf{p}_\ell, \sigma^2\mathbf{I}_L\right) \tag{2.17}$$

where $\mathbf{C} = \mathbf{\Psi}_x\mathbf{U}$ is an $N \times D$ matrix. The proposed subspace estimation procedure consists of assigning an appropriate prior distribution to $\mathbf{P}$ (denoted as $f(\mathbf{P})$) and to marginalize $\mathbf{P}$ from the joint posterior of interest.

[1] See Appendix E for further details.

It is easier to choose an informative prior distribution $f(\mathbf{P})$ that accounts for correlation between spectral bands than choosing an informative prior $f(\mathbf{W})$ since $\mathbf{P}$ is an arbitrary basis of the subspace spanned by $\mathbf{W}$, which can be easily estimated (as will be shown in the next paragraph).

2.4.3 Parameter priors

GPLVMs construct a smooth mapping from the latent space to the observation space that preserves dissimilarities (Lawrence and Quiñonero Candela, 2006). In the SU context, it means that pixels that are spectrally different have different latent variables and thus different abundance vectors. However, preserving local distances is also interesting: spectrally close pixels are expected to have similar abundance vectors and thus similar latent variables. Several approaches have been proposed to preserve similarities, including back-constraints (Lawrence and Quiñonero Candela, 2006), dynamical models (Wang and C.-I Chang, 2006) and locally linear embedding (LLE) (Urtasun et al., 2007). In this study, we use LLE to assign an appropriate prior to $\mathbf{X}$. First, the K nearest neighbors $\{\mathbf{y}_j\}_{j\in\nu_i}$ of each observation vector $\mathbf{y}_i$ are computed using the Euclidian distance (ν_i is the set of integers j such that $\mathbf{y}_j$ is a neighbor of $\mathbf{y}_i$). The weight matrix $\mathbf{\Lambda}_{\text{LLE}} = [\lambda_{i,j}]$ of size $N \times N$ providing the best reconstruction of $\mathbf{y}_i$ from its neighbors is then estimated as

$$\mathbf{\Lambda}_{\text{LLE}} = \arg\min_{\mathbf{\Lambda}} \sum_{i=1}^{N} \left\| \mathbf{y}_i - \sum_{j\in\nu_i} \lambda_{i,j}\mathbf{y}_j \right\|^2 . \tag{2.18}$$

Note that the solution of (2.18) is easy to obtain in closed form since the criterion to optimize is a quadratic function of $\mathbf{\Lambda}$. Note also that the matrix $\mathbf{\Lambda}$ is sparse since each pixel is only described by its K nearest neighbors. The locally linear patches obtained by the LLE can then be used to set the following prior for the latent variable matrix

$$f(\mathbf{X}|\mathbf{\Lambda}_{\text{LLE}}, \gamma) \propto \exp\left[-\frac{\gamma}{2} \sum_{i=1}^{N} \left\| \boldsymbol{x}_i - \sum_{j\in\nu_i} \lambda_{i,j}\boldsymbol{x}_j \right\|^2 \right] \prod_{n=1}^{N} \mathbf{1}_{\mathcal{D}}\left[\boldsymbol{x}_n\right] \tag{2.19}$$

where γ is a fixed hyperparameter to be adjusted and $\mathbf{1}_{\mathcal{D}}(\cdot)$ is the indicator function over the set $\mathcal{D}$ defined by the constraints (2.9).

In this study, we propose to assign a prior to $\mathbf{P}$ using the standard principal component analysis (PCA) (note again that the data have been centered). Assuming prior independence between $\mathbf{p}_1, \ldots, \mathbf{p}_L$, the following prior is considered for the matrix $\mathbf{P}$

$$f\left(\mathbf{P}|\overline{\mathbf{P}}, s^2\right) = \left(\frac{1}{2\pi s^2}\right)^{\frac{NL}{2}} \prod_{\ell=1}^{L} \exp\left[-\frac{1}{2s^2} \left\| \mathbf{p}_\ell - \bar{\mathbf{p}}_\ell \right\|^2 \right] \tag{2.20}$$

where $\overline{\mathbf{P}} = [\bar{\mathbf{p}}_1, \ldots, \bar{\mathbf{p}}_L]^T$ is an $L \times D$ projection matrix containing the first D eigenvectors of the sample covariance matrix of the observations (provided by PCA) and s^2 is a dispersion parameter that controls the dispersion of the prior. Note that the correlation between spectral bands is implicitly introduced through $\overline{\mathbf{P}}$. It is interesting to mention that the use of principal subspace identification methods for SU has already been investigated by Dobigeon et al. (2009a) where the projections of the endmembers onto the principal data subspace were estimated instead of the endmembers themselves.

Non-informative priors are assigned to the noise variance σ^2 and the matrix $\mathbf{U}$, i.e,

$$\begin{aligned} f(\sigma^2) &\propto \mathbf{1}_{(0,\delta_{\sigma^2})}(\sigma^2) \\ f(u_{i,j}) &\propto \mathbf{1}_{(-\delta_{\mathbf{U}},\delta_{\mathbf{U}})}(u_{i,j}) \end{aligned} \tag{2.21}$$

where the intervals $(0, \delta_{\sigma^2})$ and $(-\delta_{\mathbf{U}}, \delta_{\mathbf{U}})$ cover the possible values of the parameters σ^2 and $\mathbf{U}$. Similarly, the following non-informative prior is assigned to the hyperparameter s^2

$$f(s^2) \propto \mathbf{1}_{(0,\delta_{s^2})}(s^2) \tag{2.22}$$

where the interval $(0, \delta_{s^2})$ covers the possible values of the hyperparameter s^2. The resulting directed acyclic graph (DAG) is depicted in Fig. 2.7.

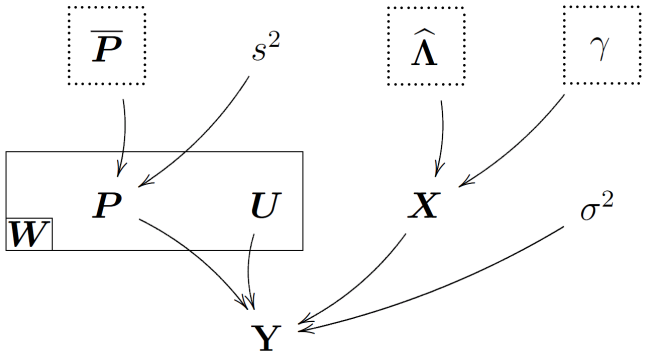


Figure 2.7: DAG for the parameter priors and hyperpriors (the fixed parameters appear in dashed boxes).

2.4.4 Marginalized posterior distribution

Assuming prior independence between $\mathbf{P}$, $\mathbf{X}$, $\mathbf{U}$, s^2 and σ^2, the marginalized posterior distribution of $\boldsymbol{\theta} = (\mathbf{X}, \mathbf{U}, s^2, \sigma^2)$ can be expressed as

$$\begin{aligned} f\left(\boldsymbol{\theta}|\mathbf{Y}, \boldsymbol{\Lambda}_{\mathrm{LLE}}, \overline{\mathbf{P}}, \gamma\right) &\propto f(\boldsymbol{\theta}|\boldsymbol{\Lambda}_{\mathrm{LLE}}, \gamma) \int f(\mathbf{Y}|\mathbf{P}, \boldsymbol{\theta}) f\left(\mathbf{P}|\overline{\mathbf{P}}, s^2\right) \mathrm{d}\mathbf{P} \\ &\propto f(\mathbf{Y}|\boldsymbol{\theta}, \overline{\mathbf{P}}) f(\boldsymbol{\theta}|\boldsymbol{\Lambda}_{\mathrm{LLE}}, \gamma) \end{aligned} \tag{2.23}$$

where $f(\boldsymbol{\theta}|\boldsymbol{\Lambda}_{\mathrm{LLE}}, \gamma) = f(\mathbf{X}|\boldsymbol{\Lambda}_{\mathrm{LLE}}, \gamma) f(\mathbf{U}) f(s^2) f(\sigma^2)$. Straightforward computations leads to

$$\begin{aligned} f(\mathbf{Y}|\boldsymbol{\theta}, \overline{\mathbf{P}}) &= \int f(\mathbf{Y}|\mathbf{P}, \boldsymbol{\theta}) f\left(\mathbf{P}|\overline{\mathbf{P}}, s^2\right) \mathrm{d}\mathbf{P} \\ &\propto \prod_{\ell=1}^{L} \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2} \bar{\mathbf{y}}_{\ell,:}^T \boldsymbol{\Sigma}^{-1} \bar{\mathbf{y}}_{\ell,:}\right] \\ &\propto |\boldsymbol{\Sigma}|^{-\frac{L}{2}} \exp\left[-\frac{1}{2} \mathrm{tr}(\boldsymbol{\Sigma}^{-1} \bar{\mathbf{Y}}^T \bar{\mathbf{Y}})\right] \end{aligned} \tag{2.24}$$

where $\boldsymbol{\Sigma} = s^2 \mathbf{C}\mathbf{C}^T + \sigma^2 \mathbf{I}_N$, $\bar{\mathbf{y}}_{\ell,:} = \mathbf{y}_{\ell,:} - \mathbf{C}\bar{\mathbf{p}}_\ell$ is an $N \times 1$ vector, $\bar{\mathbf{Y}} = [\bar{\mathbf{y}}_{1,:}, \ldots, \bar{\mathbf{y}}_{L,:}] = \mathbf{Y} - \overline{\mathbf{P}}\mathbf{C}^T$ is an $L \times N$ matrix and $\mathrm{tr}(\cdot)$ denotes the matrix trace.

Mainly due to the nonlinearity introduced through the nonlinear mapping, a closed form expression for the parameters maximizing the joint posterior distribution (2.23) is impossible to obtain. We propose to use a scaled conjugate gradient (SCG) method (Møller, 1993) to maximize the marginalized log-posterior. To ensure the sum-to-one constraint for $\mathbf{X}$, the following arbitrary reparametrization

$$x_{R,n} = 1 - \sum_{r=1}^{R-1} x_{r,n}, \quad n = 1, \ldots, N$$

is used and the marginalized posterior distribution is optimized w.r.t. the first $(R-1)$ rows of $\mathbf{X}$ denoted $\mathbf{X}_{\backslash R}$. Note that the sum-to-one constraints for the latent variables could be relaxed since we only expect the abundances of interest to satisfy these constraints. However, these constraints ensure that the dimension of the estimated manifold on which the data lie is $R-1$. The partial derivatives of the log-posterior w.r.t. $\mathbf{X}_{\backslash R}, \mathbf{U}, s^2$ and σ^2 are obtained using partial derivatives w.r.t. $\boldsymbol{\Sigma}$ and $\bar{\mathbf{Y}}$ and the classical chain rules (see Appendix F for further details). The resulting latent variable estimation procedure is referred to as locally linear GPLVM (LL-GPLVM).

Note that the marginalized likelihood reduces to the product of L independent Gaussian probability density functions since

$$\mathbf{y}_{\ell,:}|\bar{\mathbf{p}}_\ell, \mathbf{U}, \mathbf{X}, \sigma^2, s^2 \sim \mathcal{N}\left(\mathbf{C}\bar{\mathbf{p}}_\ell, s^2\mathbf{C}\mathbf{C}^T + \sigma^2\mathbf{I}_N\right) \tag{2.25}$$

and $\ell = 1, \ldots, L$. Note also that the covariance matrix $\boldsymbol{\Sigma} = s^2\mathbf{C}\mathbf{C}^T + \sigma^2\mathbf{I}_N$ is related to the covariance matrix of the 2nd order polynomial kernel (Rasmussen and Williams, 2005, p. 89). More precisely, the proposed nonlinear mapping corresponds to a particular polynomial kernel whose metric is induced by the matrix $\mathbf{U}$. Finally, note that the evaluation of the marginalized likelihood (2.24) only requires the inversion of the $N \times N$ covariance matrix $\boldsymbol{\Sigma}$. It can been seen from the following Woodbury matrix identity (Brookes, 2005)

$$\boldsymbol{\Sigma}^{-1} = \sigma^{-2}\left[\mathbf{I}_N - \mathbf{C}\left(\sigma^2 s^{-2}\mathbf{I}_D + \mathbf{C}^T\mathbf{C}\right)^{-1}\mathbf{C}^T\right] \tag{2.26}$$

that the computation of $\boldsymbol{\Sigma}^{-1}$ mainly relies on the inversion of a $D \times D$ matrix. Similarly, the computation of $|\boldsymbol{\Sigma}| = 1/|\boldsymbol{\Sigma}^{-1}|$ mainly consists of computing the determinant of a $D \times D$ matrix, which reduces the computational cost when compared to the structured covariance matrix based approach presented in paragraph 2.4.1.

2.4.5 Estimation of P

Let us denote as $\hat{\boldsymbol{\theta}} = (\widehat{\mathbf{X}}, \widehat{\mathbf{U}}, \hat{s}^2, \hat{\sigma}^2)$ the maximum a posteriori (MAP) estimator of $\boldsymbol{\theta} = (\mathbf{X}, \mathbf{U}, s^2, \sigma^2)$ obtained by maximizing (2.23). Using the likelihood (2.17), the prior distribution (2.20) and Bayes' rule, we obtain the posterior distribution of $\mathbf{P}$ conditioned upon $\boldsymbol{\theta}$, i.e.,

$$\mathbf{P}|\mathbf{Y}, \boldsymbol{\theta}, \overline{\mathbf{P}} \sim \prod_{\ell=1}^{L} \mathcal{N}\left(\mathbf{p}_\ell|\hat{\mathbf{p}}_\ell\, \mathbf{S}\right) \tag{2.27}$$

where $\mathbf{S}^{-1} = \sigma^{-2}\mathbf{C}^T\mathbf{C} + s^{-2}\mathbf{I}_D$ and $\hat{\mathbf{p}}_\ell = \mathbf{S}(\mathbf{C}^T\mathbf{y}_{\ell,:} - \bar{\mathbf{p}}_\ell)$. Since the conditional posterior distribution of $\mathbf{P}$ is the product of L independent Gaussian distributions, the MAP estimator of $\mathbf{P}$ conditioned upon $\hat{\boldsymbol{\theta}}$ is given by

$$\widehat{\mathbf{P}} = \left(\mathbf{Y}\widehat{\mathbf{C}} - \overline{\mathbf{P}}\right)\widehat{\mathbf{S}} \tag{2.28}$$

where $\widehat{\mathbf{S}}^{-1} = \hat{\sigma}^{-2}\widehat{\mathbf{C}}^T\widehat{\mathbf{C}} + \hat{s}^{-2}\mathbf{I}_D$, $\widehat{\mathbf{C}} = \widehat{\boldsymbol{\Psi}}_x\widehat{\mathbf{U}}$, $\widehat{\boldsymbol{\Psi}}_x = [\boldsymbol{\psi}_{\hat{x}}(1), \ldots, \boldsymbol{\psi}_{\hat{x}}(N)]^T$ and $\widehat{\mathbf{X}} = [\hat{\boldsymbol{x}}_1, \ldots, \hat{\boldsymbol{x}}_N]$. The MAP estimator $\widehat{\mathbf{P}}$ of $\mathbf{P}$ can be used to reconstruct the nth estimated observed pixels $\hat{\mathbf{y}}_n$ using the LL-GPLVM as follows

$$\hat{\mathbf{y}}_n = \widehat{\mathbf{P}}\widehat{\mathbf{U}}^T\boldsymbol{\psi}\left[\hat{\boldsymbol{x}}(n)\right]. \tag{2.29}$$

The next paragraph studies a scaling procedure that estimates the abundance matrix using the estimated latent variables resulting from the maximization of (2.23).

2.5 Scaling procedure

The optimization procedure presented in paragraph 2.4.4 provides a set of latent variables that represent the data but can differ from the abundance vectors of interest. Consider

$$\widehat{\mathbf{X}} = \begin{bmatrix} \widehat{\mathbf{X}}_{\backslash R} \\ \mathbf{1}_N^T - \mathbf{1}_{R-1}^T \widehat{\mathbf{X}}_{\backslash R} \end{bmatrix} \tag{2.30}$$

obtained after maximization of the posterior (2.23). The purpose of this paragraph is to estimate an $R \times N$ abundance matrix $\mathbf{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N]$ such that

$$\widehat{\mathbf{X}}_{\backslash R} = \mathbf{V}_{R-1}\mathbf{A} + \mathbf{E} \tag{2.31}$$

where $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$ occupy the maximal volume in the simplex defined by (2.8), $\mathbf{V}_{R-1} = [\mathbf{v}_1, \ldots, \mathbf{v}_R]$ is an $(R-1) \times R$ matrix and $\mathbf{E}$ is an $(R-1) \times N$ standard i.i.d Gaussian noise matrix which models the scaling errors. Since $\widehat{\mathbf{X}}$ satisfy the sum-to-one constraint (2.9), estimating the relation between $\widehat{\mathbf{X}}_{\backslash R}$ and $\mathbf{A}$ is equivalent to estimating the relation between $\widehat{\mathbf{X}}$ and $\mathbf{A}$. However, when considering the mapping between $\widehat{\mathbf{X}}$ and $\mathbf{A}$, non-isotropic noise has to be considered since the rows of $\widehat{\mathbf{X}}$ and $\mathbf{A}$ satisfy the sum-to-one constraint, i.e., they belong to the same $(R-1)$-dimensional subspace.

Eq. (2.31) corresponds to an LMM whose noisy observations are the columns of $\widehat{\mathbf{X}}_{\backslash R}$. Since $\mathbf{A}$ is assumed to occupy the largest volume in the simplex defined by (2.8), the columns of $\mathbf{V}_{R-1}$ are the vertices of the simplex of minimum volume that contains $\widehat{\mathbf{X}}_{\backslash R}$. As a consequence, it seems reasonable to use a linear unmixing strategy for the set of vectors $\widehat{\boldsymbol{x}}_{\backslash R,1}, \ldots, \widehat{\boldsymbol{x}}_{\backslash R,N}$ to estimate $\mathbf{A}$ and $\mathbf{V}_{R-1}$. In this study, we propose to estimate jointly $\mathbf{A}$ and $\mathbf{V}_{R-1}$ using the Bayesian algorithm presented in (Dobigeon et al., 2009a) for unsupervised SU assuming the LMM. Note that the algorithm in (Dobigeon et al., 2009a) assumed positivity constraints for the estimated endmembers. Since these constraints for $\mathbf{V}_{R-1}$ are unjustified, the original algorithm has slightly been modified by removing the truncations in the projected endmember priors (see (Dobigeon et al., 2009a) for details). Once the estimator $(\widehat{\mathbf{A}}, \widehat{\mathbf{V}}_{R-1})$ of $(\mathbf{A}, \mathbf{V}_{R-1})$ has been obtained by the proposed scaling procedure, we introduce constrained latent variables denoted as $\widehat{\mathbf{X}}^{(c)} = [\hat{\boldsymbol{x}}_1^{(c)}, \ldots, \hat{\boldsymbol{x}}_N^{(c)}]^T$ and defined as follows

$$\widehat{\mathbf{X}}^{(c)} = \begin{bmatrix} \widehat{\mathbf{X}}_{\backslash R}^{(c)} \\ \mathbf{1}_N^T - \mathbf{1}_{R-1}^T \widehat{\mathbf{X}}_{\backslash R}^{(c)} \end{bmatrix} \tag{2.32}$$

with $\widehat{\mathbf{X}}_{\backslash R}^{(c)} = \widehat{\mathbf{V}}_{R-1}\widehat{\mathbf{A}}$. These constrained latent variables will be used to compute the estimated observed pixels resulting from the unmixing procedure. Using the sum-to-one constraint $\widehat{\mathbf{A}}^T\mathbf{1}_R = \mathbf{1}_N$, we obtain

$$\widehat{\mathbf{X}}^{(c)} = \begin{bmatrix} \widehat{\mathbf{V}}_{R-1}\widehat{\mathbf{A}} \\ \mathbf{1}_R^T\widehat{\mathbf{A}} - \mathbf{1}_{R-1}^T\widehat{\mathbf{V}}_{R-1}\widehat{\mathbf{A}} \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{V}}_{R-1} \\ \mathbf{1}_R^T - \mathbf{1}_{R-1}^T\widehat{\mathbf{V}}_{R-1} \end{bmatrix} \widehat{\mathbf{A}} = \widehat{\mathbf{V}}_R\widehat{\mathbf{A}} \tag{2.33}$$

where $\widehat{\mathbf{V}}_R = [\widehat{\mathbf{V}}_{R-1}^T, \ \mathbf{1}_R - \widehat{\mathbf{V}}_{R-1}^T\mathbf{1}_{R-1}]^T$ is an $R \times R$ matrix. The final abundance estimation procedure, including the LL-GPLVM presented in paragraph 2.4 and the scaling procedure investigated in this paragraph is referred to as fully constrained LL-GPVLM (FCLL-GPLVM). The detailed algorithm is summarized in Algo. 2.1. The MAP estimator $\widehat{\mathbf{P}}$ in (2.28) and the estimated constrained latent variables are used to reconstruct the nth estimated observed pixels $\hat{\mathbf{y}}_n$ using the FCLL-GPLVM as follows

$$\hat{\mathbf{y}}_n = \widehat{\mathbf{P}}\widehat{\mathbf{U}}^T\boldsymbol{\psi}\left[\hat{\boldsymbol{x}}^{(c)}(n)\right] = \widehat{\mathbf{P}}\widehat{\mathbf{U}}^T\boldsymbol{\psi}\left[\widehat{\mathbf{V}}_R\hat{\boldsymbol{a}}_n\right]. \tag{2.34}$$

Once the final abundance matrix $\widehat{\mathbf{A}}$ and the matrix $\widehat{\mathbf{V}}_R$ have been estimated, we propose an endmember extraction procedure based on GP regression. This method is discussed in the next paragraph.

1: Input parameters

- Number of endmembers R.
- Matrices $\mathbf{\Lambda}_{\text{LLE}}$ and $\overline{\mathbf{P}}$.
- Number of neighbors for LLE $K = R$.
- Hyperparameter $\gamma = 10^4$.

2: Estimation

- Optimize the marginalized posterior in (2.23) using a scaled conjugate gradient algorithm to form $\hat{\boldsymbol{\theta}} = (\widehat{\mathbf{X}}, \widehat{\mathbf{U}}, \hat{s}^2, \hat{\sigma}^2)$.

3: Scaling

- Run a linear spectral unmixing algorithm on $\widehat{\mathbf{X}}_{\setminus R}$ to form the estimators $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{V}}_{(R-1)}$ (and compute the constrained latent variables according to (2.32)).

4: Endmember prediction

- Set $\boldsymbol{a} = [\mathbf{0}_{r-1}^T, 1, \mathbf{0}_{R-r}^T]^T$ and estimate the rth endmember using (2.39).

5: Output parameters

- Estimated abundance matrix $\widehat{\mathbf{A}}$.
- Estimated endmembers.

ALGO. 2.1: FCLL-GPLVM algorithm.

2.6 Gaussian process regression

Endmember estimation is one of the main issues in SU. Most of the existing EEAs intend to estimate the endmembers from the data, i.e., selecting the most pure pixels in the observed image (Chaudhry et al., 2006; Nascimento and Bioucas-Dias, 2005; Winter, 1999). However, these approaches can be inefficient when the image does not contain enough pure pixels. Some other EEAs based on the minimization of the volume containing the data (such as the minimum volume simplex analysis (Li and Bioucas-Dias, 2008)) can mitigate the absence of pure pixels in the image. This paragraph studies a new endmember estimation strategy based on GP regression for nonlinear mixtures. This strategy can be used even when the scene does not contain pure pixels. It assumes that all the image abundances have been estimated using the algorithm described in paragraph 2.5. Consider the set of pixels $\{\mathbf{y}_n\}_{n=1,\dots,N}$ and

corresponding estimated abundance vectors $\{\hat{\boldsymbol{a}}_n\}_{n=1,\ldots,N}$. GP regression first allows the nonlinear mapping $\mathbf{g}(\cdot)$ in (2.4) (from the abundance space to the observation space) to be estimated. The estimated mapping is denoted as $\hat{\mathbf{g}}(\cdot)$. Then, it is possible to use the prediction capacity of GPs to predict the spectrum $\hat{\mathbf{g}}(\boldsymbol{a})$ corresponding to any new abundance vector $\boldsymbol{a}$. In particular, the predicted spectra associated with pure pixels, i.e., the endmembers, correspond to abundance vectors that are the vertices of the simplex defined by (2.8). This paragraph provides more details about GP prediction for endmember estimation.

It can be seen from the marginalized likelihood (2.24) that $f(\mathbf{Y}|\mathbf{X}, \overline{\mathbf{P}}, \mathbf{U}, s^2, \sigma^2)$ is the product of L independent GPs associated with each spectral band of the data space (2.25). Looking carefully at the covariance matrix of $\mathbf{y}_{\ell,:}$ (i.e., $\boldsymbol{\Sigma} = s^2\mathbf{C}\mathbf{C}^T + \sigma^2\mathbf{I}_N$), we can write

$$\mathbf{y}_{\ell,:} = \mathbf{z}_\ell + \mathbf{e}_{\ell,:} \tag{2.35}$$

where $\mathbf{e}_{\ell,:}$ is the $N \times 1$ white Gaussian noise vector associated with the ℓth spectral band (having covariance matrix $\sigma^2\mathbf{I}_N$) and[2]

$$\mathbf{z}_\ell \sim \mathcal{N}\left(\mathbf{z}_\ell | \boldsymbol{\Psi}_x \mathbf{U}\bar{\mathbf{p}}_\ell, \mathbf{K}\right) \tag{2.36}$$

with $\mathbf{K} = s^2\boldsymbol{\Psi}_x\mathbf{U}\mathbf{U}^T\boldsymbol{\Psi}_x^T$ the $N \times N$ covariance matrix of $\mathbf{z}_\ell$. The $N \times 1$ vector $\mathbf{z}_\ell$ is referred to as hidden vector associated with the observation $\mathbf{y}_{\ell,:}$. Consider now an $L \times 1$ test data with hidden vector $\mathbf{z}^* = [z_1^*, ..., z_L^*]^T$, abundance vector $\boldsymbol{a}^* = [a_1^*, ..., a_R^*]^T$ and $\boldsymbol{\psi}_x^* = \boldsymbol{\psi}\left[\mathbf{V}_R\boldsymbol{a}^*\right]$. We assume that the test data share the same statistical properties as the training data $\mathbf{y}_{1,:}, ..., \mathbf{y}_{L,:}$ in the sense that $[\mathbf{z}_\ell^T, z_\ell^*]$ is a Gaussian vector such that

$$\begin{bmatrix} \mathbf{z}_\ell \\ z_\ell^* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{z}_\ell \\ z_\ell^* \end{bmatrix} \middle| \begin{bmatrix} \boldsymbol{\Psi}_x\mathbf{U}\bar{\mathbf{p}}_\ell \\ \boldsymbol{\psi}_x^{*T}\mathbf{U}\bar{\mathbf{p}}_\ell \end{bmatrix}, \begin{bmatrix} \mathbf{K} & \boldsymbol{\kappa}(\boldsymbol{a}^*) \\ \boldsymbol{\kappa}(\boldsymbol{a}^*)^T & \sigma_{\boldsymbol{a}^*}^2 \end{bmatrix} \right) \tag{2.37}$$

where $\sigma_{\boldsymbol{a}^*}^2 = s^2\boldsymbol{\psi}_x^{*T}\mathbf{U}\mathbf{U}^T\boldsymbol{\psi}_x^*$ is the variance of z_ℓ^* and $\boldsymbol{\kappa}(\boldsymbol{a}^*)$ contains the covariances between the training inputs and the test inputs, i.e.,

$$\boldsymbol{\kappa}(\boldsymbol{a}^*) = s^2\boldsymbol{\psi}_x^{*T}\mathbf{U}\mathbf{U}^T\boldsymbol{\Psi}_x. \tag{2.38}$$

Straightforward computations leads to

$$z_\ell^*|\mathbf{y}_{\ell,:} \sim \mathcal{N}\left(z_\ell^*|\mu_\ell, s_l^2\right) \tag{2.39}$$

with

$$\begin{aligned} \mu_\ell &= \boldsymbol{\psi}_x^{*T}\mathbf{U}\bar{\mathbf{p}}_\ell + \boldsymbol{\kappa}(\boldsymbol{a}^*)^T(\mathbf{K} + \sigma^2\mathbf{I}_N)^{-1}(\mathbf{y}_{:,\ell} - \boldsymbol{\Psi}_x\mathbf{U}\hat{\mathbf{p}}_\ell) \\ s_l^2 &= \sigma_{\boldsymbol{a}^*}^2 - \boldsymbol{\kappa}(\boldsymbol{a}^*)^T(\mathbf{K} + \sigma^2\mathbf{I}_N)^{-1}\boldsymbol{\kappa}(\boldsymbol{a}^*). \end{aligned}$$

Since the posterior distribution (2.39) is Gaussian, the MAP and MMSE estimators of $\mathbf{z}^*$ equal the posterior mean $\boldsymbol{\mu} = (\mu_1, ..., \mu_L)^T$.

In order to estimate the endmembers, we propose to replace the parameters $\mathbf{X}, \mathbf{U}, s^2$ and σ^2 by their estimates $\widehat{\mathbf{X}}^{(c)}, \widehat{\mathbf{U}}, \hat{s}^2$ and $\hat{\sigma}^2$ and to compute the estimated hidden vectors associated with the abundance vectors $\boldsymbol{a}^* = [\mathbf{0}_{r-1}^T, 1, \mathbf{0}_{R-r}^T]^T$ for $r = 1, \ldots, R$. For each value of r, the rth estimated hidden vector will be the rth estimated

[2] Note that all known conditional parameters have been omitted for brevity.

endmember[3]. Indeed, for the LMM and the bilinear models considered in this study, the endmembers are obtained by setting $\boldsymbol{a} = \boldsymbol{a}^* = [\mathbf{0}_{r-1}^T, 1, \mathbf{0}_{R-r}^T]^T$ in the model (2.5) relating the observations to the abundances. Note that the proposed endmember estimation procedure provides the posterior distribution of each endmember via (2.39) which can be used to derive confidence intervals for the estimates. The next paragraph presents some simulation results obtained for synthetic and real data.

2.7 Simulations on synthetic data

2.7.1 Subspace identification

The performance of the proposed GPLVM for dimensionality reduction is first evaluated on three synthetic images of $N = 2500$ pixels. The $R = 3$ endmembers contained in these images have been extracted from the spectral libraries provided with the ENVI software (RSI (Research Systems Inc.), 2003) (i.e., green grass, olive green paint and galvanized steel metal).The first image I_1 has been generated according to the linear mixing model (LMM). The second image I_2 is distributed according to the bilinear mixing model introduced in (Fan et al., 2009), referred to as the "Fan model" (FM). The third image I_3 has been generated according to the generalized bilinear model (GBM) studied in (Halimi et al., 2011a) with the following nonlinearity parameters

$$\gamma_{1,2} = 0.9, \quad \gamma_{1,3} = 0.5, \quad \gamma_{2,3} = 0.3.$$

The abundance vectors $\boldsymbol{a}_n, n = 1, \ldots, N$ have been randomly generated according to a uniform distribution on the admissible set defined by the positivity and sum-to-one constraints (2.8). The noise variance has been fixed to $\sigma^2 = 10^{-4}$, which corresponds to a signal-to-noise ratio SNR $\approx$ 30dB which corresponds to the worst case for current spectrometers. The hyperparameter γ of the latent variable prior (2.19) has been fixed to $\gamma = 10^3$ and the number of neighbors for the LLE is $K = R$ for all the results presented in this study. The quality of dimensionality reduction of the GPLVM can be measured by the ARE defined in (4.31).

Table 2.1: AREs: synthetic images ($\times 10^{-2}$).

	I_1	I_2	I_3
PCA	**0.99**	1.08	1.04
LL-GPLVM	**0.99**	**0.99**	**1.00**

Table 2.1 compares the AREs obtained by the proposed LL-GPLVM and the projection onto the first $(R-1)$ principal vectors provided by the principal component analysis (PCA). The proposed LL-GPLVM slightly outperforms PCA for nonlinear mixtures in term of ARE. More precisely, the AREs of the LL-GPLVM mainly consist of the noise errors ($\sigma^2 = 10^{-4}$), whereas model errors are added when applying PCA to nonlinear mixtures. Fig. 2.8 compares the latent variables obtained after maximization of (2.24) for the three images I_1 to I_3 with the projections obtained by projecting the data onto the $R-1$ principal vectors provided by PCA. Note that only $R-1$ dimensions are needed to represent the latent variables (because of the sum-to-one constraint). From this figure, it can be seen that

[3]Note that the estimated endmembers are centered since the data have previously been centered. The actual endmembers can be obtained by adding the empirical mean of the data to the estimated endmembers.

the latent variables of the LL-GPLVM describe a noisy simplex for the three images. It is not the case when using PCA for the nonlinear images. Fig. 2.9 shows the manifolds estimated by the LL-GPLVM for the three images I_1 to I_3. This figure shows that the proposed LL-GPLVM can model the manifolds associated with the image pixels with good accuracy.

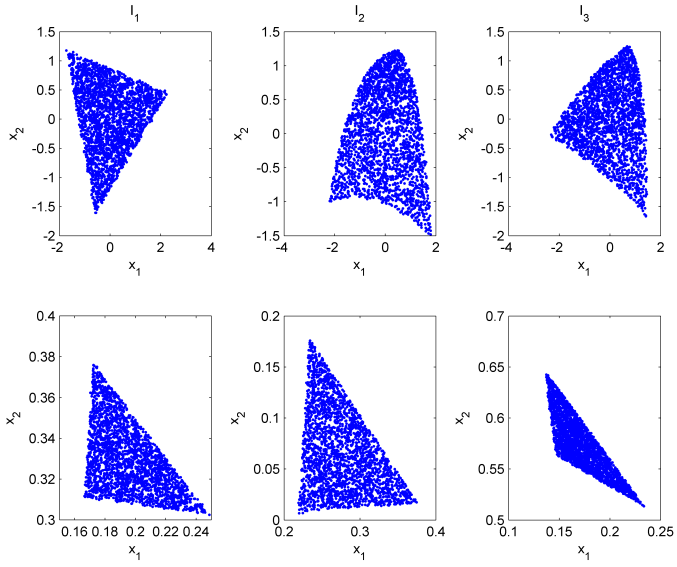


Figure 2.8: Top: Representation of the $N = 2500$ pixels (dots) using the first two principal components provided by the standard PCA for the three synthetic images I_1 to I_3. Bottom: Representation using the latent variables estimated by the LL-GPLVM for the three synthetic images I_1 to I_3.

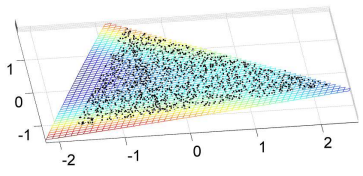
2.7.2 Abundance and endmember estimation

The quality of unmixing procedures can be measured by comparing the estimated and actual abundances using the RNMSE defined in (1.36). Table 2.2 compares the RNMSEs obtained with different unmixing strategies. The endmembers have been estimated by the VCA algorithm in all simulations. The algorithms used for abundance estimation are the FCLS algorithm proposed in (Heinz and C.-I Chang, 2001) for I_1, the LS method proposed in (Fan et al., 2009) for I_2 and the gradient-based method proposed in (Halimi et al., 2011a) for I_3. These procedures are referred to as "SU" in the table. These strategies are compared with the proposed FCLL-GPLVM. As mentioned above, the Bayesian algorithm for joint estimation of $\mathbf{A}$ and $\mathbf{V}$ under positivity and sum-to-one constraints for $\mathbf{A}$ (introduced by Dobigeon et al. (2009a)) is used in this paragraph for the scaling step. It can be seen that the proposed FCLL-GPLVM is general enough to accurately approximate the considered mixing models since it provides the best results in term of abundance estimation.

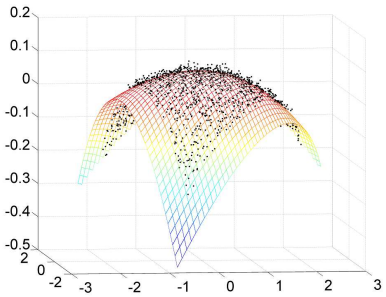
Table 2.2: RNMSEs: synthetic images ($\times 10^{-3}$).

	I_1	I_2	I_3
SU	5.7	7.4	22.7
FCLL-GPLVM	**3.9**	**4.2**	**5.4**

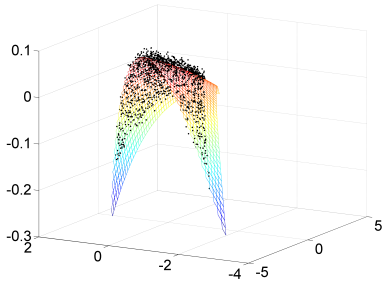
The quality of reconstruction of the unmixing procedure is also evaluated by the ARE. Table 2.3 shows the AREs



(a) I_1 (LMM)



(b) I_2 (FM)



(c) I_3 (GBM)

Figure 2.9: Visualization of the $N = 2500$ pixels (black dots) of I_1, I_2 and I_3 using the 3 axis provided by the PCA procedure. The colored surface is the manifold identified by the LL-GPLVM.

corresponding to the different unmixing strategies. The proposed FCLL-GPLVM outperforms the other strategies in term of ARE for these images.

Table 2.3: AREs: synthetic images ($\times 10^{-2}$).

	I_1	I_2	I_3
SU	1.00	1.13	1.06
FCLL-GPLVM	**0.99**	**0.99**	**1.00**

Finally, the performance of the FCLL-GPLVM for endmember estimation is evaluated by comparing the estimated endmembers with the actual spectra. The quality of endmember estimation is evaluated by the SAM defined in (1.71). Table 2.4 compares the SAMs obtained for each endmember using the VCA algorithm, the nonlinear EEA presented in (Heylen et al., 2011) (referred to as "Heylen") and the FCLL-GPLVM for the three images I_1 to I_3. These results show that the FCLL-GPLVM provides accurate endmember estimates for both linear and nonlinear mixtures.

2.7.3 Performance in absence of pure pixels

The performance of the proposed unmixing algorithm is also tested in scenarios where pure pixels are not present in the observed scene. More precisely, the simulation parameters remain the same for the three images I_1 to I_3

Table 2.4: SAMs ($\times 10^{-2}$): synthetic images.

	I_1			I_2			I_3		
	$\mathbf{m}_1$	$\mathbf{m}_2$	$\mathbf{m}_3$	$\mathbf{m}_1$	$\mathbf{m}_2$	$\mathbf{m}_3$	$\mathbf{m}_1$	$\mathbf{m}_2$	$\mathbf{m}_3$
VCA	**0.43**	**0.22**	0.22	1.62	2.08	1.15	1.91	1.36	0.88
Heylen	1.94	0.66	0.78	0.75	1.69	0.42	1.80	0.86	1.38
FCLL-GPLVM	0.52	0.86	**0.15**	**0.33**	**0.53**	**0.34**	**0.44**	**0.58**	**0.30**

except for the $N = 2500$ abundance vectors, that are drawn from a uniform distribution in the following set

$$\left\{ \boldsymbol{a} | \sum_{r=1}^{R} a_r = 1, \quad 0.9 \geq a_r(n) \geq 0, \forall r \in \{1, \ldots, R\} \right\}. \tag{2.40}$$

The three resulting images are denoted as I_1^*, I_2^* and I_3^*. Table 2.5 shows that the absence of pure pixels does not significantly change the AREs of the FCLL-GPLVM when they are compared with those obtained with the images I_1 to I_3 (see Tables 2.1 and 2.3). Moreover, FCLL-GPLVM is more robust to the absence of pure pixels than the different SU methods. The good performance of FCLL-GPVLM is due in part to the scaling procedure. Table 2.6 shows that the performance of the FCLL-GPLVM in term of RNMSE is not degraded significantly when there is no pure pixel in the image (see Table 2.2 for comparison). Note that the situation is different when the endmembers are estimated using VCA. Table 2.7 shows the performance of the FCLL-GPLVM for endmember estimation when there is no pure pixel in the image. The results of the FCLL-GPLVM do not change significantly when they are compared with those obtained with images I_1 to I_3, which is not the case for the two other EEAs. The accuracy of the endmember estimation is illustrated in Fig. 2.10 which compares the endmembers estimated by the FCLL-GPLVM (blue lines) to the actual endmember (red dots) and the VCA estimates (black line) for the image I_2^*. Similar endmember estimates have been obtained for the images I_1^* and I_3^* (same components involved in the mixtures).

Table 2.5: AREs: synthetic images (absence of pure pixels, $\times 10^{-2}$).

	PCA	LL-GPLVM	SU	FCLL-GPLVM
I_1^*	**1.00**	**1.00**	1.14	**1.00**
I_2^*	1.06	**1.00**	1.57	**1.00**
I_3^*	1.03	**0.99**	1.12	**0.99**

Table 2.6: RNMSEs: synthetic images (absence of pure pixels, $\times 10^{-3}$).

	I_1^*	I_2^*	I_3^*
SU	49.3	86.6	47.8
FCLL-GPLVM	**4.8**	**7.2**	**7.5**

2.7.4 Performance with respect to endmember variability

The proposed method assumes that the spectrum characterizing a given material (i.e., an endmember) is unique for all the image pixels. This assumption has been widely used in linear unmixing, which has motivated the consideration

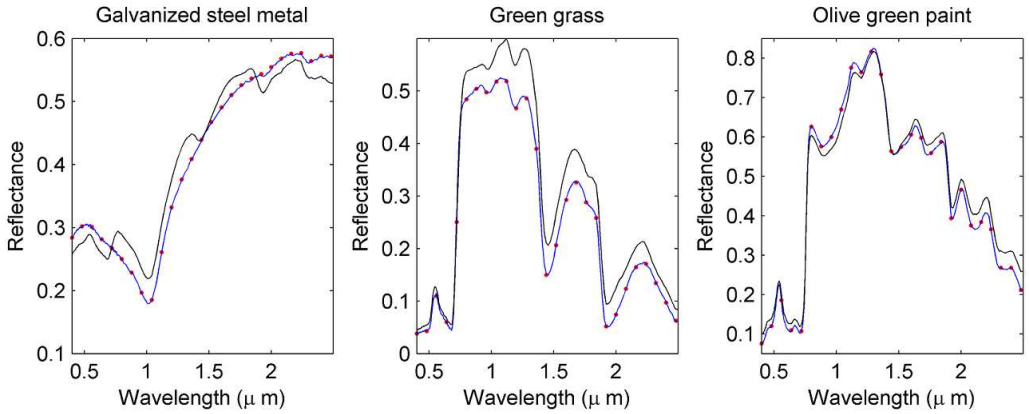


Figure 2.10: Actual endmembers (red dots) and endmembers estimated by the FCLL-GPLVM (blue lines) and VCA (black line) for the image I_2^*.

Table 2.7: SAMs ($\times 10^{-2}$): synthetic images.

	I_1^*			I_2^*			I_3^*		
	$\mathbf{m}_1$	$\mathbf{m}_2$	$\mathbf{m}_3$	$\mathbf{m}_1$	$\mathbf{m}_2$	$\mathbf{m}_3$	$\mathbf{m}_1$	$\mathbf{m}_2$	$\mathbf{m}_3$
VCA	2.87	2.15	2.10	5.22	8.02	7.10	6.89	6.03	3.73
Heylen	6.38	11.11	2.62	7.53	9.59	2.48	6.59	5.95	2.36
FCLL-GPLVM	**0.38**	**1.30**	**0.24**	**0.67**	**1.46**	**0.53**	**0.61**	**1.75**	**0.48**

of unique endmembers. However, taking endmember variability into consideration is also an important problem, depending on the observation conditions and the observed scene (Eches et al., 2010; Somers et al., 2011; Zare et al., 2012). To evaluate the robustness of the proposed method to endmember variability, additional experiments have been performed. More precisely, sets of $N = 2500$ synthetic pixels have been generated according to the following nonlinear model

$$\mathbf{y}(n) = \sum_{r=1}^{R} \mathbf{m}_r(n) a_r(n) + \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} a_i(n) a_j(n) \mathbf{m}_i(n) \odot \mathbf{m}_j(n) + \mathbf{e}(n),$$

where $\boldsymbol{a}(n) = [a_1(n), \ldots, a_R(n)]^T$ has been generated uniformly in the simplex defined by the positivity and sum-to-one constraints and endmember variability has been considered by using random endmembers, i.e., $\mathbf{m}_r(n) \sim \mathcal{N}\left(\mathbf{m}_r^0, \sigma_{\mathbf{M}}^2 \mathbf{I}\right)$ where $\mathbf{m}_r^0, r = 1, \ldots, R$ are the actual endmembers extracted from the spectral library and $\sigma_{\mathbf{M}}^2$ is the endmember variance. Note that this model is similar to the Fan model studied by Fan et al. (2009) except the fact that the endmembers are random. Table 2.8 compares the performance of the proposed method with the performance of an unmixing strategy based on VCA (for endmember extraction) and the least squares method by Fan et al. (2009) (for abundance estimation). This procedure is referred to as "SU" in the table. Four values of $\sigma_{\mathbf{M}}^2$ have been considered. The higher $\sigma_{\mathbf{M}}^2$, the higher the endmember variability. For each row, the best result has been highlighted in blue. The spectral angle mappers (SAMs) presented in Table 2.8 represent the angles between the estimated endmembers and the actual endmembers $\mathbf{m}_r^0, r = 1, \ldots, R$. From this table, it can be seen that for each value of $\sigma_{\mathbf{M}}^2$, the proposed method provides more accurate abundance and endmember estimates (in term of

RNMSE and SAM, respectively), when compared with the SU approach. In particular, the performance of the proposed method is not significantly degraded for weak endmember variability.

Table 2.8: Endmember variability: synthetic images.

			$\sigma^2_{\mathbf{M}} = 0$	$\sigma^2_{\mathbf{M}} = 10^{-5}$	$\sigma^2_{\mathbf{M}} = 10^{-4}$	$\sigma^2_{\mathbf{M}} = 10^{-3}$
RNMSE ($\times 10^{-3}$)	SU		7.4	9.1	11.0	23.2
	FCLL-GPLVM		**4.2**	**5.1**	**9.1**	**11.9**
SAM ($\times 10^{-2}$)	SU	$\mathbf{m}_1$	1.62	2.06	2.09	2.55
		$\mathbf{m}_2$	2.08	1.67	**1.94**	2.38
		$\mathbf{m}_3$	1.15	1.27	1.06	1.23
	FCLL-GPLVM	$\mathbf{m}_1$	**0.33**	**0.44**	**0.46**	**1.59**
		$\mathbf{m}_2$	**0.53**	**0.71**	2.01	**1.62**
		$\mathbf{m}_3$	**0.34**	**0.47**	**0.42**	**0.99**

The next paragraph presents simulation results obtained for real data.

2.8 Application to a real dataset

The real image considered in this paragraph is a part of the image acquired over Villelongue, France and introduced in Chapter 1. This dataset has already been studied by Sheeren et al. (2011) and is composed of a forested area containing 12 identified vegetation species (ash tree, oak tree, hazel tree, locust tree, chestnut tree, lime tree, maple tree, beech tree, birch tree, willow tree, walnut tree and fern). The sub-image of size 50×50 pixels chosen here to evaluate the proposed unmixing procedure is depicted in Fig. 4.6. A reasonably small image is considered here to ease the explanation of the results and to keep the processing overhead quite low.

This image contains vegetation species of varying spatial density such that some pixels do not contain identified tree species. More precisely, the scene is mainly composed of three components since the data belong to a two-dimensional manifold (see black dots of Fig. 2.12 (a)). Consequently, we assume that the scene is composed of $R = 3$ endmembers. We propose to use the set of 32224 label spectra used by Sheeren et al. (2011) for the learning step of the classification method presented herein to identify the components present in the area of interest. More precisely, Fig. 2.12 (a) shows the reference clusters corresponding to oak trees (red dots) and chestnut trees (blue dots) projected in a 3-dimensional subspace (defined by the first three principal components of a PCA applied to the image of Fig. 4.6). These two clusters are the two closest sets of pixels to vertices of the data cloud. Consequently, oak and chestnut trees are identified as endmembers present in the image. Moreover, the new identified endmember is associated with the non-vegetation area (the strategy conducted in (Sheeren et al., 2011) was restricted to vegetation species). In the sequel, this endmember will be referred to as Endmember ♯3.

The simulation parameters have been fixed to $\gamma = 10^3$ and $K = R$. The latent variables obtained by maximizing the marginalized posterior distribution (2.14) are depicted in Fig. 2.13 (blue dots).

It can be seen from this figure that the latent variables seem to describe a noisy simplex. Fig. 2.12 (b) shows the manifold estimated by the proposed LL-GPLVM. This figure illustrates the capacity of the LL-GPLVM for modeling the nonlinear manifold. Table 2.9 (left) compares the AREs obtained by the proposed LL-GPLVM and

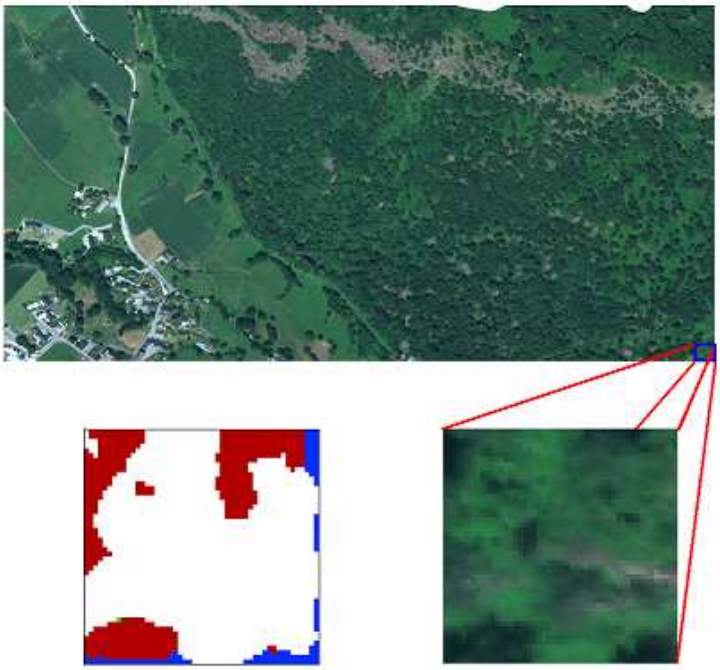


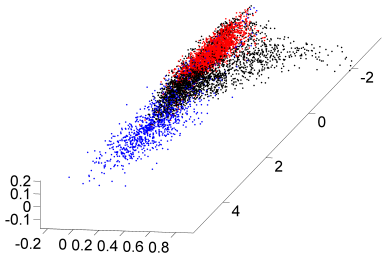
Figure 2.11: Top: real hyperspectral Madonna data acquired by the Hyspex hyperspectral scanner over Villelongue, France. Bottom right: the region of interest shown in true colors (right). Bottom left: classification map obtained in (Sheeren et al., 2011) for the region of interest. The labeled pixels are classified as Oak tree (red), Chestnut tree (blue), Ash tree (green) and non-planted-tree pixels (white).

the projection onto the first $R - 1 = 2$ principal vectors provided by PCA. The proposed LL-GPLVM slightly outperforms PCA for the real data of interest, which shows that the proposed nonlinear dimensionality reduction method is more accurate than PCA (linear dimensionality reduction) in representing the data.

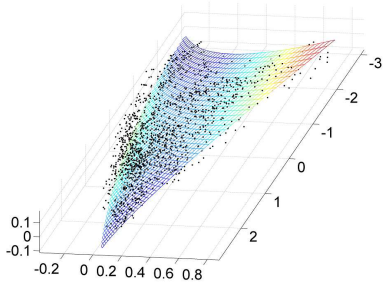
Table 2.9: AREs: real image ($\times 10^{-2}$).

PCA	LL-GPLVM	VCA+FCLS	FCLL-GPLVM
0.84	**0.79**	1.30	**1.11**

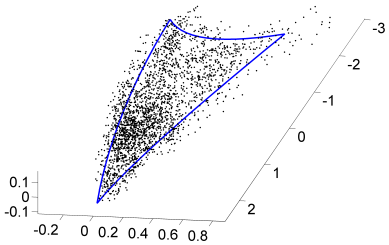
The scaling step presented in paragraph 2.5 is then applied to the estimated latent variables. The estimated simplex defined by the latent variables is depicted in Fig. 2.13 (red lines). Fig. 2.12 (c) compares the boundaries of the estimated transformed simplex with the image pixels. The abundance maps obtained after the scaling step are shown in Fig. 2.14 (top). The results of the unmixing procedure using the FCLL-GPLVM are compared to an unmixing strategy assuming the LMM. More precisely, we use VCA to extract the endmembers from the data and use the FCLS algorithm for abundance estimation. The estimated abundance maps are depicted in Fig. 2.14 (bottom). The abundance maps obtained by the two methods are similar which shows the accuracy of the proposed unmixing strategy when considering the LMM as a first order approximation of the mixing model. Moreover, Fig. 4.6 (bottom left) shows the classification map obtained by Sheeren et al. (2011) for the region of interest.
The white pixels correspond to areas where the classification method by Sheeren et al. (2011) has not been performed. Since the aim of the work presented by Sheeren et al. (2011) was to locate tree species, a non-planted-tree reference



(a)



(b)



(c)

Figure 2.12: (a): Representation of the $N = 2500$ pixels (black dots) of the Villelongue image and the reference clusters corresponding to oak trees (red dots) and chestnut trees (blue dots) using the first three principal components provided by the standard PCA. (b): Representation of the $N = 2500$ pixels (dots) of the Villelongue data and manifold identified by the LL-GPLVM (colored surface). (c):Representation of the $N = 2500$ pixels (dots) of the Villelongue data and boundaries of the estimated transformed simplex (blue lines).

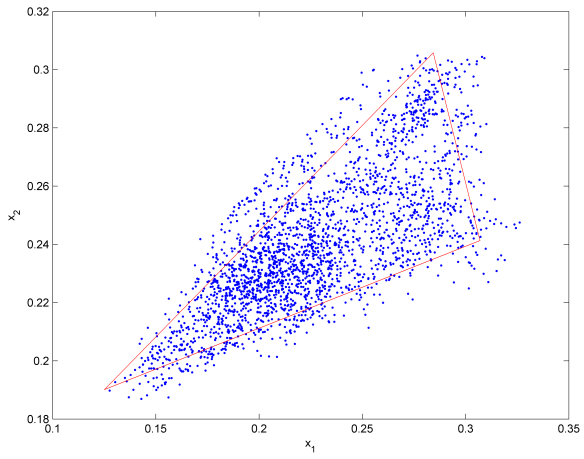


Figure 2.13: Representation of the $N = 2500$ latent variables (dots) estimated by the LL-GPLVM and the simplex identified by the scaling step (red lines) for the Villelongue data.

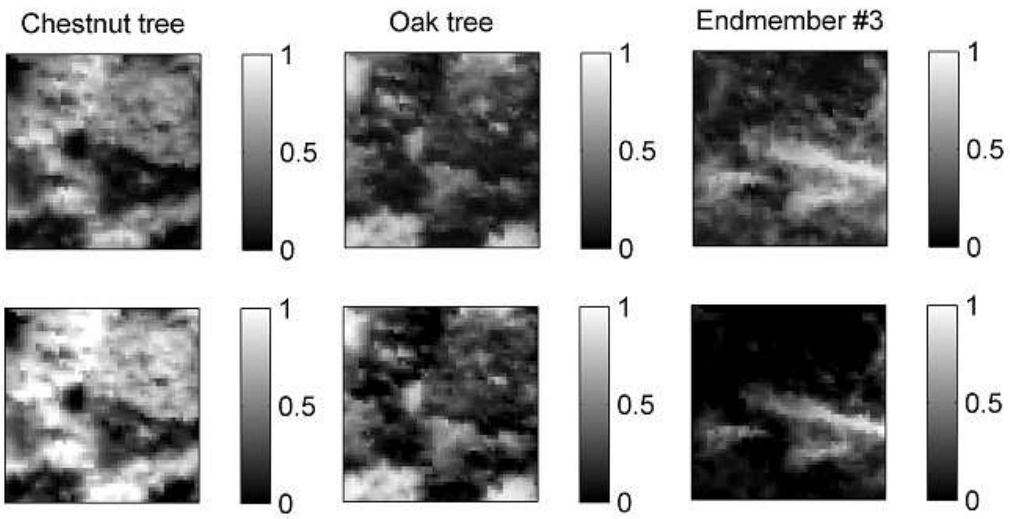


Figure 2.14: Top: Abundance maps estimated using the FCLL-GPLVM for the Villelongue image. Bottom: Abundance maps estimated using the VCA algorithm for endmember extraction and the FCLS algorithm for abundance estimation.

mask was used to classify only planted-tree pixels. Even if lots of pixels are not classified, the classified pixels can be compared with the estimated abundance maps. First, we can note the presence of the same tree species in the classification and abundance maps, i.e., oak and chestnut. We can also see that the pixels composed of chestnut trees and Endmember ♯3 are mainly located in the unclassified regions, which explains why they do not appear clearly in the classification map. Only one pixel is classified as being composed of ash trees in the region of interest. If unclassified pixels also contain ash trees, they are either too few or too mixed to be considered as mixtures of an additional endmember in the image. Finally, it can be seen from Figs. 2.14 and 4.6 (bottom left) that oak trees are located within similar regions (left corners and top right corner) for the abundance and classification maps.

Evaluating the performance of endmember estimation on real data is an interesting problem. However, comparison of the estimated endmembers with the ground truth is difficult here. First, since the nature of Endmember ♯3 is unknown, no ground truth is available for this endmember. Second, because of the variability of the ground truth spectra associated with each tree species, it is difficult to show whether VCA or the proposed FCLL-GPLVM provides the best endmember estimates. However, the AREs obtained for both methods (Table 2.9, right) show that the FCLL-GPLVM fits the data better than the linear SU strategy, which confirms the importance of the proposed algorithm for nonlinear spectral unmixing.

In this chapter, the number of pure spectral components is assumed to be known. When using linear mixing models, algorithms such as the HySime algorithm (Bioucas-Dias and Nascimento, 2008) are able to identify the number of endmembers in the scene, based on the estimation of the data subspace. However, there is no equivalent algorithm developed for images defined by nonlinear mixtures. Of course, if we assume that the model nonlinearities are weak (compared to the linear part), the HySime algorithm can be used to approximate the number of endmembers. For the considered real image, the dimension of the subspace spanned by the data estimated by HySime is $R = 6$. This value can be used to upperbound the number of endmembers since the estimated signal subspace possibly includes nonlinearities.

Table 2.10: Estimation of R: real image.

	ARE ($\times 10^{-2}$)			
	$R=2$	$R=3$	$R=4$	$R=5$
LL-GPLVM	1.58	0.79	0.74	0.71
FCLL-GPLVM	1.58	1.11	1.08	0.86

Another possibility is to run the proposed algorithm for different values of R and compare the results. The proposed unmixing procedure has been applied to the studied real image for different values of R (i.e., $R = 2, 3, 4, 5$). Table 2.10 compares the AREs of the LL-GPLVM and the FCLL-GPLVM for this image. From this table, it can be seen that the higher R, the lower the AREs for the dimensionality reduction method and the FCLL-GPLVM algorithm, as expected. However, the decrease of ARE for the LL-GPLVM is less significant when the number of components has been overestimated. Conversely, the value of ARE is increasing significantly when the number of endmembers has been underestimated. This strategy can be used for estimating the value of R. Note that a similar method was proposed by Quirion et al. (2008) for estimating the dimension of the latent variable subspace.

Finally, Fig. 2.15 shows the abundance maps estimated by the FCLL-GPLVM (on the real image) for different values of R (i.e., $R = 2, 3, 4, 5$). From this figure, it can be seen that the two vegetation species (i.e., chestnut tree and oak tree) are merged when R is underestimated, mainly due to the high correlation between these two components. When the number of endmembers is overestimated, the locations of oak trees and of the third endmember are estimated similarly. However, the abundance maps of the chestnut tree are less structured for $R = 4$ and $R = 5$ than for $R = 3$. It can also be observed that the abundances maps of the fourth and fifth endmembers do not correspond to homogeneous regions. The analysis of the estimated endmembers for $R = 5$ shows that the estimated endmembers $2, 4$ and 5 are similar (see Fig. 2.16). These considerations could be used to make sure the value of R has been designed correctly.

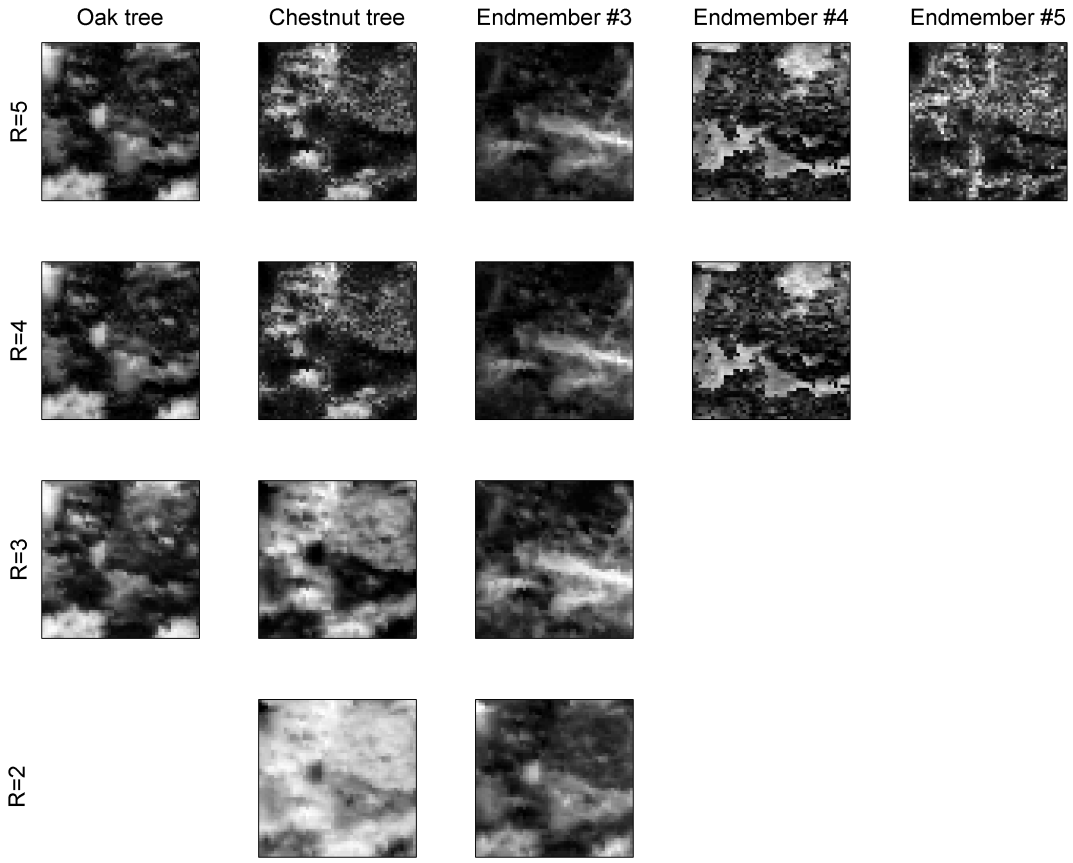


Figure 2.15: Abundance maps estimated using the FCLL-GPLVM for the Madonna image for $R = 2$ (bottom) to $R = 5$ (top).

2.9 Conclusion

We proposed a new algorithm for nonlinear spectral unmixing based on a Gaussian process latent variable model. The unmixing procedure assumed a nonlinear mapping from the abundance space to the observed pixels. It also considered the physical constraints for the abundance vectors. The abundance estimation was decomposed into two steps. Dimensionality reduction was first achieved using latent variables. A scaling procedure was then proposed to estimate the abundances. After estimating the abundance vectors of the image, a new endmember estimator based on Gaussian process regression was investigated. This decomposition of the unmixing procedure, consisting of first estimating the abundance vectors and subsequently the endmembers, breaks the usual paradigm of spectral unmixing. Simulations conducted on synthetic data illustrated the flexibility of the proposed model for linear and nonlinear spectral unmixing and provided promising results for abundance and endmember estimations even when there are few pure pixels in the image. It was shown in this chapter that the proposed unmixing procedure provides better or comparable performance (in terms of abundance and endmember estimation) than state of the

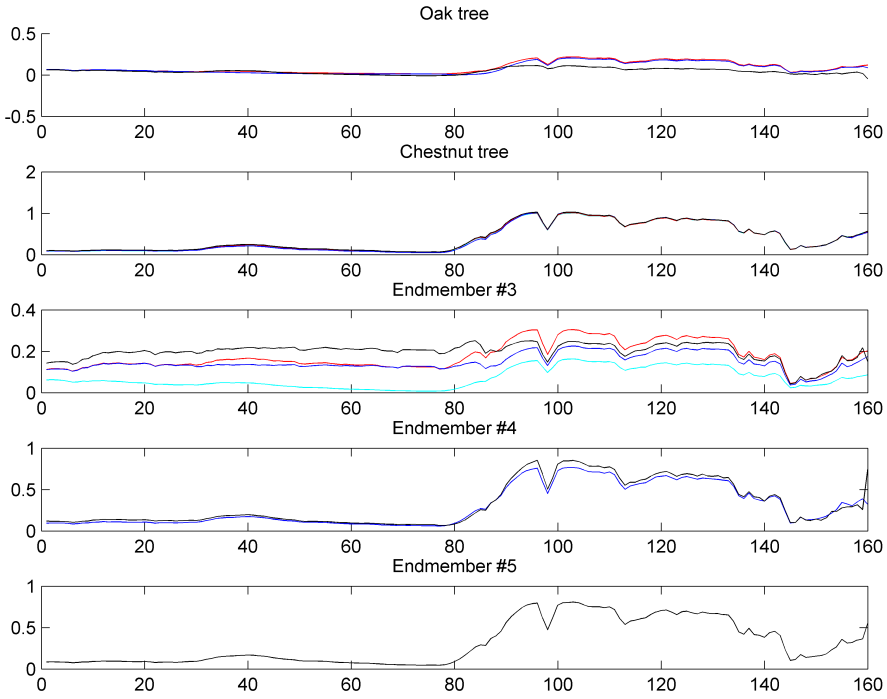


Figure 2.16: Endmembers of the real image estimated by the FCLL-GPLVM for $R = 2$ (light blue), $R = 3$ (red), $R = 4$ (blue) and $R = 5$ (black).

art unmixing strategies assuming specific mixing models.

The proposed abundance estimation procedure presented in this chapter was split into two steps. Thus, the abundance estimates (and consequently the predicted endmembers) mainly rely on the scaling step. It could be interesting to assign more accurate priors for the latent variables to get rid of that scaling step. In that case, the latent variables would be the abundances. Additional spectral information was included within the LL-GPLVM by assuming that spectrally closed pixels should have closed latent variables. From a strict Bayesian point of view, such information should be included in the observation model, i.e., in the likelihood, to set prior distributions independent of the observations. The nonlinear dimensionality reduction presented in this chapter relies on the maximization of an appropriate posterior distribution using a gradient-based method. To avoid possible convergence issues, such as being trapped in local maxima and setting a stopping rule for the iterative process, it could be interesting to couple the proposed Bayesian model with efficient simulation methods such as Markov chain Monte Carlo methods.

Finally, the choice of the nonlinear mapping used for the GP model is an important issue to ensure that the LL-GPLVM is general enough to handle different nonlinearities. In particular, different mappings could be used for intimate mixtures. However, the tradeoff between model flexibility and estimation performance is difficult to find, especially when addressing the unsupervised unmixing problem. Fitting the data manifold can be accurately achieved using different kernel functions. However, the associated latent variables may be not obviously related to the abundances.

The proposed FCLL-GPVLM assumed the observed pixels resulting from mixtures of R materials to belong to an $(R-1)$-dimensional manifold. Of course this assumption holds only for a single mixing process fully parameterized

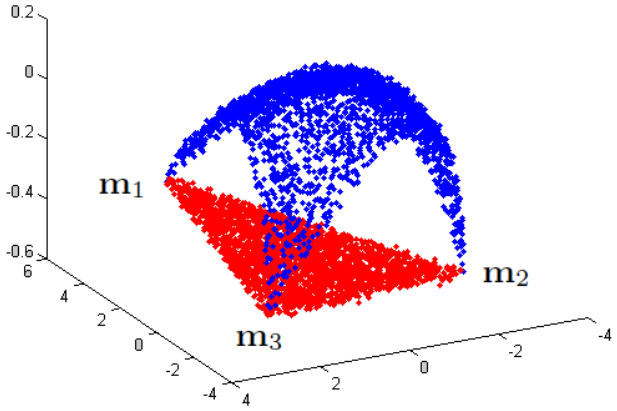


Figure 2.17: Representation of mixed pixels composed of $R = 3$ endmembers, distributed according to the LMM (red) and the bilinear FM (blue).

by the abundance vectors. If different kinds of mixtures occur in the same image, the FCLL-GPVLM would try to estimate a single manifold while several manifolds could be present, as depicted in Fig. 2.17. For such situations, other unmixing methods should be used. To address the problem of multiple mixing processes in hyperspectral, it seems natural to identify image regions where different kinds of mixtures are present. Unmixing strategies assuming a single model could then be used in each region independently, depending on the nature of these regions. The next chapter focuses on binary classifiers to identify linearly and nonlinearly mixed pixels in hyperspectral images.

Main contribution. A new nonparametric nonlinear SU algorithm was proposed for unsupervised SU. The proposed unmixing algorithm breaks the usual paradigm of spectral unmixing by first estimating the abundances and then predicting the endmembers even in the absence of pure pixel in the image.

2.10 Conclusion (in French)

Dans ce chapitre, nous avons proposé un nouvel algorithme pour le démélange spectral non-linéaire à l'aide d'un modèle à variables latentes et de processus gaussiens. La procédure de démélange suppose une relation non-linéaire entre l'espace des abondances et celui des observations. Les contraintes physiques qui restreignent les abondances ont également été prises en compte. L'estimation des abondances est décomposée en deux étapes. La première consiste en une étape de réduction de dimension faisant intervenir des variables latentes. Une procédure de mise à l'échelle est alors proposée pour estimer les abondances des composants de l'image pour chaque pixel. Suite à l'estimation des abondances, les spectres des composants purs de l'image sont estimés par régression à base de processus gaussiens. Cette décomposition de la méthode de démélange, constituée de l'estimation des abondances avant l'estimation des signatures spectrales, casse le paradigme habituel du démélange spectral. Les simulations effectuées sur données synthétiques ont illustré la flexibilité du modèle proposé pour le démélange linéaire et non-linéaire. Des résultats prometteurs en termes d'estimation des signatures spectrales ont été obtenus, même s'il y a peu ou pas de pixels purs dans l'image. Il a été montré dans ce chapitre que cette méthode de démélange permet d'obtenir de meilleures performances (en termes d'estimation des abondances et des signatures spectrales) qu'avec des algorithmes récents supposant des modèles de mélange plus spécifiques.

La procédure d'estimation des abondances présentée dans ce chapitre a été divisée en deux étapes. Les abondances estimées (et par conséquent les spectres estimés) reposent principalement sur l'étape de mise à l'échelle. Il serait intéressant de considérer de l'information *a priori* plus discriminante pour les variables latentes afin de s'affranchir de cette étape de mise à l'échelle. Dans ce cas, les variables latentes seraient directement les abondances. De l'informations spectrale supplémentaire a été incluse dans le LL- GPLVM, en supposant que les pixels spectralement proches ont des variables latentes proches. D'un point de vue strictement bayésien, cette information devrait être incluse dans le modèle d'observation, c'est-à-dire dans la vraisemblance, pour définir des lois *a priori* indépendantes des observations. L'étape de réduction de dimension non-linéaire présentée dans ce chapitre repose sur la maximisation d'une loi *a posteriori* d'intérêt en utilisant une méthode de gradient. Pour éviter d'éventuels problèmes de convergence, il pourrait être intéressant de coupler le modèle bayésien proposé à des méthodes de simulation efficaces telles que des méthodes MCMC.

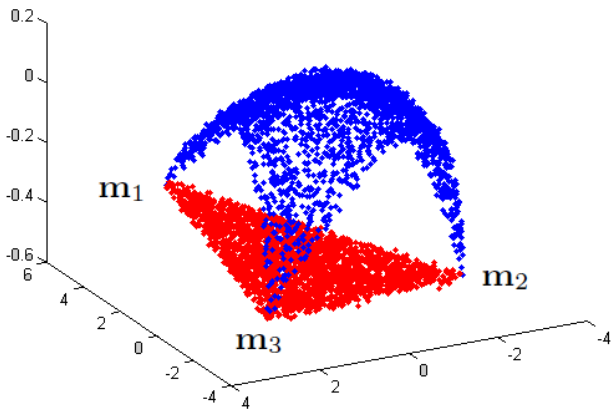


Figure 2.18: Representation de pixels résultant de mélanges de $R = 3$ composants générés suivant le modèle de mélange linéaire (rouge) et le modèle bilinéaire de Fan (bleu).

Enfin, le choix de la relation non-linéaire utilisée pour le modèle proposé est un point important pour assurer un LL-GPLVM assez général. En particulier, des noyaux différents pourraient être utilisés pour des mélanges dits intimes. Toutefois, le compromis entre la flexibilité du modèle et les performances en termes d'estimation des abondances et des signatures spectrales est difficile à trouver, surtout lors de la résolution du problème de démélange non-supervisé. L'approximation de la variété sur laquelle vivent les données peut se faire à l'aide de noyaux différents. Cependant, la relation entre les variables latentes et les abondances peut ne pas être triviale selon le noyau utilisé.

L'algorithme FCLL-GPVLM proposé suppose que les pixels observés résultant de mélanges de R composants appartiennent à une variété de dimension $(R-1)$. Bien sûr, cette hypothèse ne vaut que pour un processus de mélange unique entièrement paramétré par les vecteurs d'abondances. Si différents types de mélanges avaient lieu dans une même image, le FCLL-GPVLM essaierait d'identifier une seule variété alors que plusieurs seraient présentes, comme le montre la Fig. 2.18. Pour de telles situations, d'autres méthodes de démélange doivent être utilisées. Pour résoudre le problème de plusieurs types de mélanges dans une même image, il semble naturel d'identifier les zones où les differents types de mélanges sont présents. Des algorithmes de démélange (basés sur des modèles uniques) pourraient ensuite être utilisés dans chaque région de façon indépendante, en fonction de la nature de ces régions. Le chapitre suivant étudie deux classifieurs binaires pour identifier les pixels résultant de mélanges linéaires et non-linéaires dans les images hyperspectrales.

Contributions majeures. Un nouvel algorithme non-paramétrique a été proposé pour le démélange non-linéaire non-supervisé. Cet algorithme brise le paradigme habituel du démélange spectral en estimant d'abord les abondances et en prédisant ensuite les signatures spectrales des composants purs, même s'il n'y a pas de composants purs dans l'image.

Chapter 3

Nonlinearity detection in hyperspectral images

The first part of this chapter has been adapted from the journal paper (Altmann et al., 2012) (published).

Contents

3.1 Introduction (in French)

Ce chapitre aborde le problème de détection de mélanges non-linéaires dans les images hyperspectrales. On cherche donc à déterminer si un pixel observé résulte d'un mélange linéaire ou non-linéaire des signatures spectrales des composants purs présents dans l'image. Dans ce chapitre, les composants de l'image (nombre et spectres) sont supposés connus. Les problèmes liés à l'estimation des composants de l'image précédant la détection de non-linéarités seront discutés dans la conclusion de ce chapitre.

Le premier détecteur de non-linéarités proposé dans ce chapitre est basé sur le modèle PPNMM étudié pour le démélange non-linéaire dans le chapitre 1. Une des propriétés intéressantes du modèle PPNMM est qu'il généralise le modèle de mélange linéaire grâce à un paramètre de non-linéarité unique dont la valeur caractérise la non-linéarité dans le pixel considéré. En particulier, lorsque le paramètre de non-linéarité est égal à zéro, le modèle de mélange résultant est linéaire. Par conséquent, il semble naturel d'utiliser ce paramètre pour construire un détecteur de non-linéarités.

Le second détecteur de non-linéarité étudié dans ce chapitre repose uniquement sur le modèle de mélange linéaire. Il ne suppose donc aucune caractérisation particulière de la non-linéarité, ce qui le rend plus robuste aux divers types de non-linéarités. Le test statistique associé à ce détecteur consiste à calculer la distance de chaque pixel observé à l'hyperplan défini par les spectres des composants purs de l'image et la contrainte de somme-à-un des abondances. Cette distance est ensuite comparée à un seuil (liée au niveau de bruit) au delà duquel on décide que le modèle linéaire classique n'est plus capable de modéliser précisement le mélange. Les performances des deux détecteurs sont analysées et comparées à l'aide de simulations menées sur données synthétiques et réelles.

3.2 Introduction

This chapter addresses the problem of determining whether an observed pixel of a hyperspectral image results from a linear or nonlinear mixture of endmembers. In this chapter, the endmembers (number and spectra) are assumed to be known. Endmember estimation issues for nonlinearity detection will be discussed in the conclusion of this chapter. The first proposed nonlinearity detector is based on the PPNMM studied for nonlinear SU in Chapter 1. One of the most interesting properties of the PPNMM is that it generalizes the LMM thanks to a unique nonlinearity parameter whose value characterizes the nonlinearity in the considered pixel. In particular, when the nonlinearity parameter equals zero, the resulting mixing model is linear. Consequently, it seems natural to use this parameter for deriving new nonlinearity detectors. The second nonlinearity detector studied in this chapter only relies on the LMM, i.e., does not assume any particular nonlinearity characterization which makes it more robust to various kinds of nonlinearities.

3.3 Supervised PPNMM-based nonlinearity detection

This paragraph recalls the polynomial post-nonlinear mixing model introduced in Chapter 1 for nonlinear unmixing and used in this chapter for nonlinearity detection in hyperspectral images. The associated parameter estimation algorithm is also briefly summarized. The statistical test for nonlinearity detection based on the parameter estimators provided by the unmixing procedure is then derived. Constrained Cramér-Rao lower bounds (CCRLBs) of the PPNMM parameter estimators are finally used to approximate the variance of the nonlinearity parameter estimator yielding an approximated test statistics for nonlinearity detection.

3.3.1 PPNMM model and parameter estimation

According to the PPNMM introduced in Chapter 1 the L-spectrum $\mathbf{y} = [y_1, \ldots, y_L]^T$ of a mixed pixel is defined as a polynomial post-nonlinear transformation of a linear mixture of R endmembers $\mathbf{m}_1, \ldots, \mathbf{m}_R$ contaminated by additive noise

$$\mathbf{y} = \mathbf{M}\boldsymbol{a} + b(\mathbf{M}\boldsymbol{a}) \odot (\mathbf{M}\boldsymbol{a}) + \mathbf{e}. \tag{3.1}$$

The noise sequence $\mathbf{e}$ is an additive independent and identically distributed (i.i.d) zero-mean Gaussian noise sequence with variance σ^2, denoted as $\mathbf{e} \sim \mathcal{N}\left(\mathbf{0}_L, \sigma^2 \mathbf{I}_L\right)$. In this paragraph, the endmember matrix $\mathbf{M}$ is assumed to be known. Consequently, the only unknown parameters are the abundance vector $\boldsymbol{a}$, the nonlinearity parameter b and the noise variance σ^2.

The abundance vector $\boldsymbol{a}$ satisfies the following positivity and sum-to-one constraints

$$\sum_{r=1}^{R} a_r = 1, \quad a_r \geq 0, \forall r \in \{1, \ldots, R\}. \tag{3.2}$$

In Chapter 1, it has been shown that the PPNMM is general enough to handle a wide class of nonlinear models, which is interesting for nonlinearity detection. Moreover, it has also been shown that the PPNMM parameters can be estimated by minimizing the following LS criterion

$$\begin{aligned} J(\boldsymbol{a}, b) &= \frac{1}{2} \left\| \mathbf{y} - \mathbf{g}(\mathbf{M}\boldsymbol{a}) \right\|^2 \\ &= \frac{1}{2} \left\| \mathbf{y} - \mathbf{M}\boldsymbol{a} - b(\mathbf{M}\boldsymbol{a}) \odot (\mathbf{M}\boldsymbol{a}) \right\|^2 \end{aligned} \tag{3.3}$$

subject to the constraints (3.2). After estimating $\boldsymbol{a}$ and b, the noise variance σ^2 can be estimated as follows

$$\hat{\sigma}^2 = \frac{1}{L}\left\|\mathbf{y} - \mathbf{M}\hat{\boldsymbol{a}} - \hat{b}(\mathbf{M}\hat{\boldsymbol{a}}) \odot (\mathbf{M}\hat{\boldsymbol{a}})\right\|^2. \tag{3.4}$$

Since the additive noise vector $\mathbf{e}$ is assumed to be an i.i.d zero-mean Gaussian noise, the resulting estimator of $\boldsymbol{\theta} = [\boldsymbol{a}^T, b, \sigma^2]^T$ is the maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$, denoted as $\hat{\boldsymbol{\theta}}$. Consequently, the estimator $\hat{\boldsymbol{\theta}} = [\hat{\boldsymbol{a}}^T, \hat{b}, \hat{\sigma}^2]^T$ is asymptotically efficient and asymptotically distributed according to a Gaussian distribution (Kay, 1993, Chap. 7) . Note that the asymptotic region corresponds here to $L \to \infty$. Since L is very large (some hundreds of spectral bands) for hyperspectral images, the asymptotic region will be reached in most practical applications[1]. The two LS algorithms considered in Chapter 1 (i.e., based on linearization and subgradient methods) for minimizing (3.3) subject to the constraints (3.2) have provided very similar performance. As a consequence, this study will concentrate on one estimator only, namely the subgradient-based estimator. The Bayesian algorithm introduced in paragraph 1.4.1 could also be considered. However, it should be slightly modified by assigning uniform prior distributions over suitable sets for all the unknown parameters, such that the MAP and MLE estimators coincide. The next paragraph derives a nonlinearity detector based on the MLE of the nonlinearity parameter.

3.3.2 Nonlinearity detection

As shown in Chapter 1, the PPNMM allows the nonlinearity to be characterized by the parameter b for each pixel of the scene. An arbitrary threshold could be used to decide if the observed pixel is better modeled by the LMM or by a general nonlinear model defined by (3.1). However, it would be difficult to choose the appropriate threshold in order to guarantee a given probability of false alarm (PFA) or a given probability of detection (PD). In this paragraph, we propose a statistical test for pixel-by-pixel nonlinearity detection based on the distribution of $\hat{b}$. Based on the asymptotic properties of the MLE and on the large number of spectral bands available for a hyperspectral image, it makes sense to approximate the distribution of $\hat{b}$ by the following Gaussian distribution[2]

$$\hat{b} \sim \mathcal{N}\left(b, s^2\right) \tag{3.5}$$

where $s^2 \triangleq s^2(\boldsymbol{a}, b, \sigma^2)$ is the variance of the estimator $\hat{b}$. It is important to note that the variance of $\hat{b}$ is a function of the parameters $\boldsymbol{a}$, b and σ^2. Obviously, when the observation vector $\mathbf{y}$ results from the LMM (i.e., $b = 0$), then

$$\hat{b} \sim \mathcal{N}\left(0, s_0^2\right) \tag{3.6}$$

where $s_0^2 = s^2(\boldsymbol{a}, 0, \sigma^2)$. This interesting property can be used for testing the mixing model appropriate to the observation vector. The resulting nonlinearity detection problem can be considered as a two hypothesis testing problem, where the hypotheses are defined as

$$\begin{cases} H_0 & : \quad \mathbf{y} \text{ is distributed according to the LMM} \\ H_1 & : \quad \mathbf{y} \text{ is distributed according to the PPNMM.} \end{cases} \tag{3.7}$$

Hypothesis H_0 is characterized by $b = 0$ whereas nonlinear models (H_1) correspond to $b \neq 0$. As a consequence, the two hypotheses in (3.7) can be rewritten as

$$\begin{cases} H_0 & : \quad \hat{b} \sim \mathcal{N}\left(0, s_0^2\right) \\ H_1 & : \quad \hat{b} \sim \mathcal{N}\left(b, s_1^2\right) \end{cases} \tag{3.8}$$

where $s_1^2 = s^2(\boldsymbol{a}, b, \sigma^2)$ and $b \neq 0$. Estimating the variance $s^2(\boldsymbol{a}, b, \sigma^2)$ will be discussed in paragraph 3.3.3.

[1] The asymptotic behavior of the considered MLEs will be discussed in paragraph 3.3.4.

[2] This assumption will be validated in the simulation results.

Known parameters $\boldsymbol{a}$ and σ^2

For a given observation vector $\mathbf{y}$ and its corresponding estimated nonlinearity parameter $\hat{b}$, we propose to decide between hypotheses H_0 and H_1 using a classical generalized likelihood ratio test (GLRT) for (3.8). Using (3.5) and (3.6), the probability density functions of the test statistic $\hat{b}$ under the two hypothesis can be written

$$p(\hat{b}|H_0) = \left(\frac{1}{2\pi s_0^2}\right)^{\frac{1}{2}} \exp\left(-\frac{\hat{b}^2}{2s_0^2}\right) \tag{3.9}$$

$$p(\hat{b}|H_1) = \left(\frac{1}{2\pi s_1^2}\right)^{\frac{1}{2}} \exp\left(-\frac{(\hat{b}-b)^2}{2s_1^2}\right). \tag{3.10}$$

The corresponding GLRT consists of comparing the test statistic

$$\frac{\sup\limits_{b} \; p(\hat{b}|H_1)}{p(\hat{b}|H_0)} \tag{3.11}$$

to an appropriate threshold. Obviously, $p(\hat{b}|H_1)$ is maximized for $b = \hat{b}$. Straightforward computations lead to the following test strategy

$$T = \frac{\hat{b}^2}{s_0^2} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{3.12}$$

where η is a threshold that is related to the test PFA as follows

$$\begin{aligned} P_{\text{FA}} &= \mathbb{P}\left[\frac{\hat{b}^2}{s_0^2} > \eta \middle| H_0\right] \\ &= 2\mathbb{P}\left[\frac{\hat{b}}{s_0} < -\sqrt{\eta} \middle| H_0\right] = 2\phi(-\sqrt{\eta}) \end{aligned} \tag{3.13}$$

where $\phi(\cdot)$ is the cumulative distribution function of the normalized Gaussian distribution. It is interesting to note that the test (3.12) does not depend on s_1^2 and only relies on the variance s_0^2 under hypothesis H_0. For a given value of b, the power of the test $P_{\text{D}}(b)$ can be computed as follows

$$P_{\text{D}}(b) = \mathbb{P}\left[\frac{\hat{b}^2}{s_0^2} > \eta \middle| H_1\right] = \mathbb{P}\left[\frac{\hat{b}^2}{s_0^2} > \eta \middle| b \neq 0\right]. \tag{3.14}$$

Straightforward computations lead to

$$\begin{aligned} P_{\text{D}}(b) &= \mathbb{P}\left[\frac{\hat{b}-b}{s_1} < \frac{-s_0\sqrt{\eta}-b}{s_1} \middle| H_1\right] \\ &+ \mathbb{P}\left[\frac{\hat{b}-b}{s_1} > \frac{s_0\sqrt{\eta}-b}{s_1} \middle| H_1\right] \\ &= 1 + \phi\left(\frac{-s_0\sqrt{\eta}-b}{s_1}\right) - \phi\left(\frac{s_0\sqrt{\eta}-b}{s_1}\right). \end{aligned} \tag{3.15}$$

It can be observed that for a given value of the threshold η, the probability of detection $P_{\text{D}}(b)$ is an increasing function of $|b|$, which is an intuitive result. In order to apply the detection strategy (3.12) and to compute the corresponding P_{FA} and $P_{\text{D}}(b)$, we need to know the parameters s_0 and s_1 whose determination is the objective of the next paragraph.

Unknown parameters $\boldsymbol{a}$ and σ^2

The test (3.12) assumes known parameters $\boldsymbol{a}$ and σ^2 to compute $s_0^2 = s^2(\boldsymbol{a}, 0, \sigma^2)$. However, these parameters are unknown in practical applications. To alleviate this problem, we propose to approximate the variance of $\hat{b}$ under H_0 by an appropriate estimator $\hat{s}_0^2$ leading to

$$\hat{T} = \frac{\hat{b}^2}{\hat{s}_0^2} \underset{H_0}{\overset{H_1}{\gtrless}} \eta^*. \tag{3.16}$$

More precisely, in order to build $\hat{s}_0^2$, we propose to use the constrained Cramér-Rao lower bound (CRLB) of $\boldsymbol{\theta} = [\boldsymbol{a}^T, b, \sigma^2]^T$ under hypothesis H_0 (i.e., $b = 0$) as explained in the next paragraph.

3.3.3 Constrained Cramér-Rao bound

This paragraph studies the constrained Cramér-Rao lower bound associated with any unbiased estimator $\hat{\boldsymbol{\theta}}$ of the parameter vector $\boldsymbol{\theta}$ involved in the PPNMM. Eq. (3.1) shows that $\mathbf{y}|\boldsymbol{a}, b, \sigma^2 \sim \mathcal{N}\left(\mathbf{g}\left(\mathbf{M}\boldsymbol{a}\right), \sigma^2\mathbf{I}_L\right)$. As a consequence, the likelihood function of $\mathbf{y}$ is defined as

$$f(\mathbf{y}|\boldsymbol{a}, b, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{L}{2}} \exp\left(-\frac{\|\mathbf{y} - \mathbf{g}\left(\mathbf{M}\boldsymbol{a}\right)\|^2}{2\sigma^2}\right). \tag{3.17}$$

The corresponding unconstrained CRLB for any unbiased estimator of $\boldsymbol{\theta}$ constructed from $\mathbf{y}$ is given by

$$\text{CRLB}(\boldsymbol{\theta}) = \mathbf{J}_F^{-1} \tag{3.18}$$

where $\mathbf{J}_F$ is the Fisher information matrix whose elements are[3]

$$[\mathbf{J}_F]_{i,j} = -\text{E}_{\mathbf{y}|\boldsymbol{\theta}}\left[\frac{\partial^2 \ln f(\mathbf{y}|\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}\right] \quad i, j = 1, \ldots, R+2.$$

However, the positivity and sum-to-one constraints (3.2) are not considered in this expression. Particularly, the sum-to-one constraint in (3.2) enforces the R-dimensional abundance vector $\boldsymbol{a}$ to belong to an $(R-1)$-dimensional subspace. This constraint can be considered by computing a reduced-rank Fisher matrix yielding a constrained Cramér-Rao lower bound (CCRLB). The CCRLB principles have been introduced by Gorman and Hero (1990) for parameters satisfying equality and/or inequality constraints. The constraints for the abundance vector in (3.2) can be rewritten

$$u_{\boldsymbol{\theta}} = \begin{bmatrix}\mathbf{1}_R^T & 0 & 0\end{bmatrix} \begin{bmatrix}\boldsymbol{a} \\ b \\ \sigma^2\end{bmatrix} - 1 = \mathbf{c}^T\boldsymbol{\theta} - 1 = 0 \tag{3.19}$$

and

$$\mathbf{v}_{\boldsymbol{\theta}} = \begin{bmatrix}-\mathbf{I}_R & \mathbf{0}_R & \mathbf{0}_R\end{bmatrix} \begin{bmatrix}\boldsymbol{a} \\ b \\ \sigma^2\end{bmatrix} = \mathbf{A}\boldsymbol{\theta} \preceq \mathbf{0}_R \tag{3.20}$$

where $\mathbf{1}_R$ is an $R \times 1$ vector of ones, $\mathbf{c} = [\mathbf{1}_R^T, 0, 0]^T$ is an $(R+2) \times 1$ vector, $\mathbf{A}$ is an $R \times (R+2)$ matrix, $u_{\boldsymbol{\theta}}$ is the equality constraint, $\mathbf{v}_{\boldsymbol{\theta}}$ is an $R \times 1$ pure inequality vector (see (Gorman and Hero, 1990) for details) and $\preceq$ denotes the termwise inequality. Since the set of admissible $\boldsymbol{\theta}$ is an $(R+1)$-dimensional subset of $\mathbb{R}^{R+2}$, the CCRLB

[3] The Fisher information matrix $\mathbf{J}_F$ is derived in Appendix G.

associated with the covariance matrix of any constrained unbiased estimator of $\boldsymbol{\theta}$ is given by (Gorman and Hero, 1990)

$$\text{CCRLB}(\boldsymbol{\theta}) = \mathbf{Q}\mathbf{J}_F^{-1} \tag{3.21}$$

with

$$\mathbf{Q} \quad = \quad \mathbf{I}_{R+2} - \mathbf{J}_F^{-1}\nabla u_{\boldsymbol{\theta}} \left\{\nabla u_{\boldsymbol{\theta}}^T \mathbf{J}_F^{-1} \nabla u_{\boldsymbol{\theta}}\right\} \nabla u_{\boldsymbol{\theta}}^T$$

where, from (3.19), $\nabla u_{\boldsymbol{\theta}} = \mathbf{c}$ is the gradient of $u_{\boldsymbol{\theta}}$. It is interesting to note that the CCRLB can be easily computed since this matrix results from simple operations applied on the unconstraint CRLB $\mathbf{J}_F^{-1}$ and the vector $\mathbf{c}$. Moreover, no arbitrary reparametrization of the problem is needed. The CCRLB of b is then given by the $(R+1)$th diagonal element of $\text{CCRLB}(\boldsymbol{\theta})$ denoted as $\text{CCRLB}(b; \boldsymbol{a}, \sigma^2)$. An estimator of the variance of $\hat{b}$ under hypothesis H_0 is required to compute the test statistic (3.16). In this study, we propose to estimate s_0^2 as follows

$$\hat{s}_0^2 = \text{CCRLB}(0; \hat{\boldsymbol{a}}, \hat{\sigma}^2) \tag{3.22}$$

where $\hat{\boldsymbol{a}}$ and $\hat{\sigma}^2$ are the MLEs of $\boldsymbol{a}$ and σ^2. The next paragraphs study the performance of the nonlinearity detector defined by (3.16), where $\hat{s}_0^2$ is defined in (3.22), for synthetic and real hyperspectral data.

3.3.4 Synthetic data

Estimation

The statistical test proposed in (3.16) assumes the efficiency and normality of the estimator $\hat{b}$ resulting from the unmixing procedure. We first propose to show that the asymptotic region in term of MLE efficiency is usually reached in the hyperspectral imagery context (i.e., for large L and high signal to-noise ratio (SNR)). Four different mixtures are considered to illustrate the estimator efficiency. These mixtures are composed of $R = 3$ materials (i.e., green grass, olive green paint and galvanized steel metal) whose endmember spectra, composed of $L = 826$ bands, have been extracted from the spectral libraries provided with the ENVI software (RSI (Research Systems Inc.), 2003). The synthetic mixtures have been obtained using the following parameters[4]

$$\begin{aligned}
&\text{Mixture} \quad \mathcal{M}_0 \rightarrow \boldsymbol{a} = [0.3, 0.6, 0.1]^T, b = 0, \sigma^2 = 3 \times 10^{-3} \\
&\text{Mixture} \quad \mathcal{M}_1 \rightarrow \boldsymbol{a} = [0.5, 0.1, 0.4]^T, b = 0, \sigma^2 = 3 \times 10^{-3} \\
&\text{Mixture} \quad \mathcal{M}_2 \rightarrow \boldsymbol{a} = [0.3, 0.6, 0.1]^T, b = 0.2, \sigma^2 = 3 \times 10^{-3} \\
&\text{Mixture} \quad \mathcal{M}_3 \rightarrow \boldsymbol{a} = [0.3, 0.6, 0.1]^T, b = 0, \sigma^2 = 1 \times 10^{-3}.
\end{aligned}$$

The efficiency of the proposed unmixing algorithm is evaluated by comparing the CCRLB with the mean square errors (MSEs)

$$\text{MSE}(\hat{b}) = \frac{1}{N}\sum_{n=1}^{N}\left[\hat{b}_n - b_n\right]^2, \quad i = 1, \ldots, R+2 \tag{3.23}$$

associated with the nonlinearity parameter b, where N is the number of pixels to be unmixed and $\hat{b}_n$ is the estimated value of the nth actual parameter b_n .

[4] Note that $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$ have been obtained by changing $\boldsymbol{a}$, b and σ^2 in $\mathcal{M}_0$, respectively.

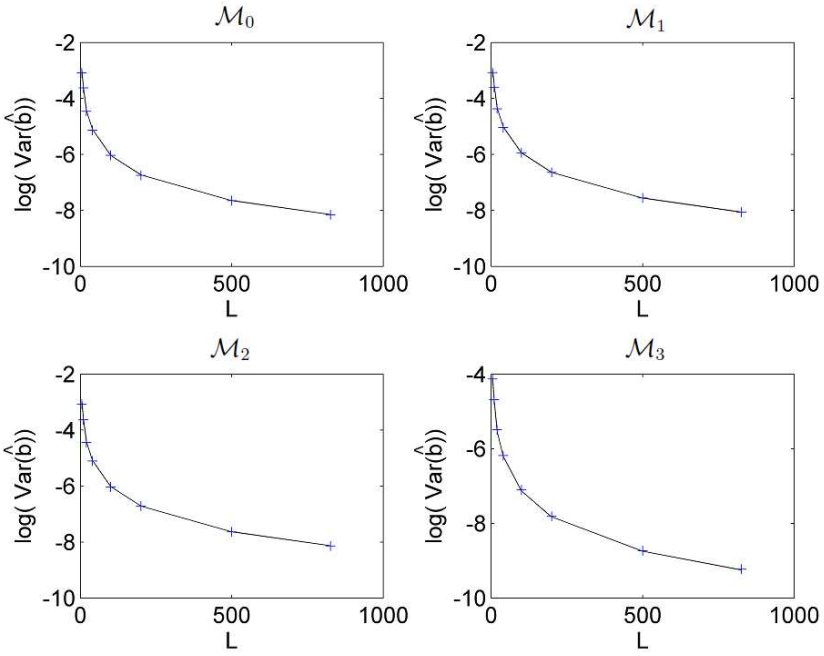


Figure 3.1: MSEs of the MLE (blue crosses) for the nonlinearity parameter b versus L compared with the CCRLBs (black lines).

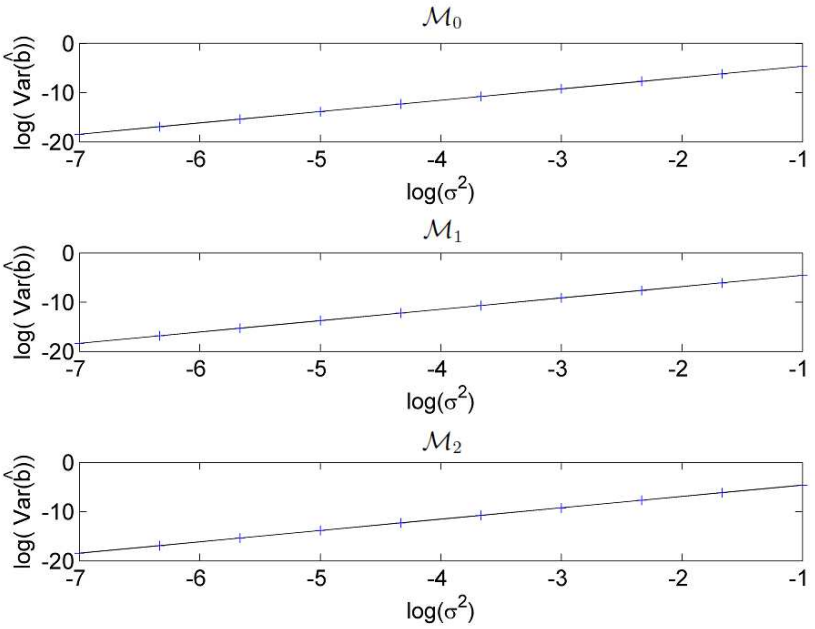


Figure 3.2: MSEs of the MLE (blue crosses) for the nonlinearity parameter b versus σ^2 compared with the CCRLBs (black lines).

Fig. 3.1 compares the MSEs of the subgradient-based estimator[5], estimated with $N = 20000$ noise realizations,

[5]Similar results have been obtained using the linearization-based estimator and are reported in (Altmann et al., 2011).

with the CCRLB versus the number of spectral bands (the number of spectral bands has been adjusted by a regular subsampling of the initial $L = 826$ bands). These results confirm the efficiency of the MLE for these four mixtures since the MSEs (crosses) are very close to the corresponding CCRLBs (continuous lines).
Similarly, Fig. 3.2 compares the MSEs of the MLE (estimated with $N = 20000$ noise realizations and $L = 826$) with the CCRLB versus the noise variance σ^2 for the mixed pixels $\mathcal{M}_0$ to $\mathcal{M}_2$. These results show that the efficiency property is valid for any value of σ^2.

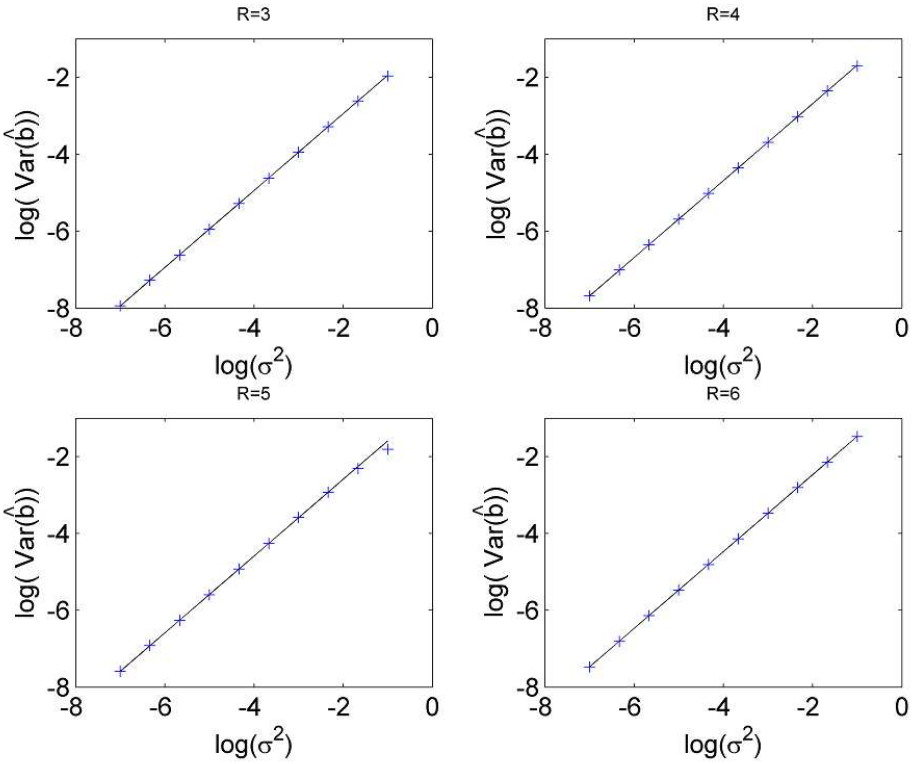


Figure 3.3: MSEs of the MLE (blue crosses) for the nonlinearity parameter b and $R = 3, 4, 5, 6$ versus σ^2 compared with the CCRLBs (black lines).

Fig. 3.3 compares the MSEs of the MLE (estimated with $N = 20000$ noise realizations and $L = 826$) with the CCRLB versus the noise variance σ^2 for $R = 3, 4, 5, 6$. The considered endmembers are the three materials presented above and construction concrete, micaceous loam and bare red brick. The synthetic mixtures have been obtained using the following parameters

$$\begin{aligned} R = 3 : \boldsymbol{a} &= [0.3, 0.6, 0.1]^T, \quad b = 0.2 \\ R = 4 : \boldsymbol{a} &= [0.2, 0.3, 0.3, 0.2]^T, \quad b = 0.1 \\ R = 5 : \boldsymbol{a} &= [0.1, 0.15, 0.15, 0.2, 0.4]^T, \quad b = -0.2 \\ R = 6 : \boldsymbol{a} &= [0.15, 0.20, 0.25, 0.1, 0.1, 0.2]^T, \quad b = -0.1 \end{aligned}$$

These results show that the efficiency assumption of the MLE of b is valid for different values of R. The asymptotic normality for the MLE of b is then investigated by considering the distributions of $\hat{b}$ for the four mixtures $\mathcal{M}_0$ to $\mathcal{M}_3$. The histograms of $\hat{b}$ estimated from $N = 20000$ Monte Carlo runs are depicted in Fig. 3.4. Moreover, single sample

Kolmogorov-Smirnov goodness-of-fit tests Kay (1998) with significance level set to 0.05 have been applied to the four sets of Monte Carlo runs (corresponding to $\mathcal{M}_0$ to $\mathcal{M}_3$). The four tests have accepted the null hypothesis (i.e., the distributions of $\hat{b}$ coincide with their asymptotic distributions) with large $p-$values $(0.93; 0.79; 0.58$ and 0.72 for $\mathcal{M}_0$ to $\mathcal{M}_3)$. These results confirm that the distributions of the subgradient-based algorithm can be approximated by a Gaussian distribution whose mean is the actual parameter b and whose variance is given by the CCRLB.

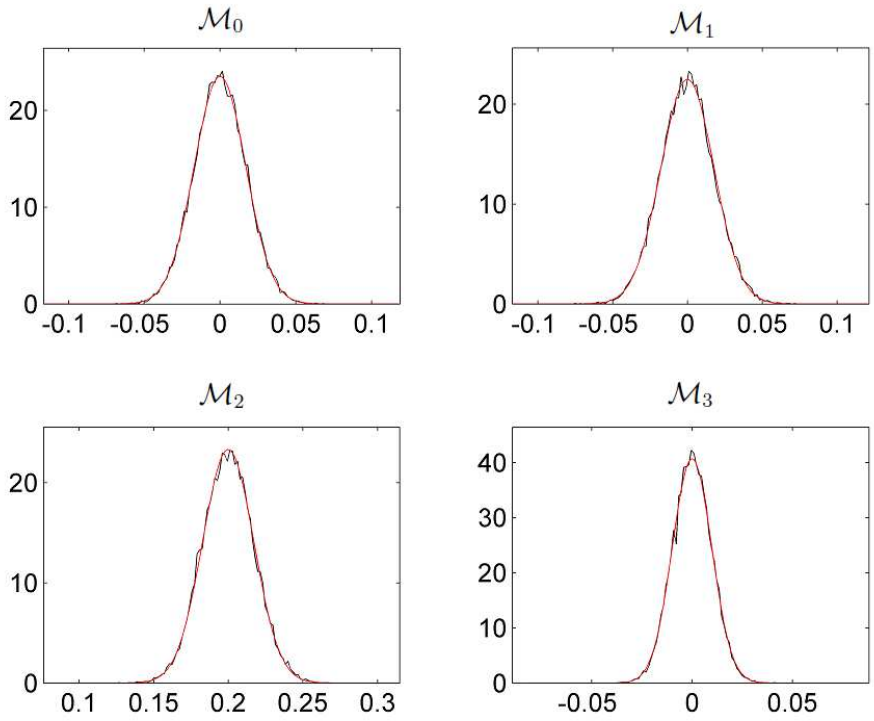


Figure 3.4: Histograms of $\hat{b}$ (black lines) and associated Gaussian distributions (red lines) for the four mixtures $\mathcal{M}_0$ to $\mathcal{M}_3$.

Detection performance The performance of the proposed nonlinearity detection procedure can be measured by comparing the actual PFA (given by (3.13)) with the empirical PFA defined as

$$P_{\text{FA}}^{\text{emp}}(\eta) = \frac{1}{N}\sum_{n=1}^{N} d_n(\eta) \tag{3.24}$$

with

$$d_n(\eta) = \begin{cases} 0, \text{if} & T_n < \eta \\ 1, \text{if} & T_n > \eta \end{cases} \tag{3.25}$$

where N is the number of noisy realizations of a given mixture under H_0, η is the theoretical test threshold, T_n is the value of the test statistic for the nth noise realization and $d_n(\cdot)$ is its corresponding decision ($d_n(\cdot) = i$ means hypothesis H_i has been accepted with $i \in \{0, 1\}$). The actual PFA is also compared to its approximation obtained

by approximating the CCRLB

$$P_{\text{FA}}^{\text{app}}(\eta) = \frac{1}{N}\sum_{n=1}^{N} \hat{d}_n(\eta) \tag{3.26}$$

where

$$\hat{d}_n(\eta) = \begin{cases} 0, \text{if} \quad \hat{T}_n < \eta \\ 1, \text{if} \quad \hat{T}_n > \eta. \end{cases} \tag{3.27}$$

Fig. 3.5 compares P_{FA}, $P_{\text{FA}}^{\text{emp}}$ and $P_{\text{FA}}^{\text{appr}}$ as a function of the threshold η for $N = 20000$ noisy realizations of the mixture $\mathcal{M}_0$. These results first show that the theoretical and empirical PFAs coincide. Moreover, the CCRLB approximation proposed for the final test does not modify the performance in term of PFA. Fig. 3.6 shows the test performance in term of receiver operating characteristics (ROCs) (Kay, 1998, p. 74-75) for $\boldsymbol{a} = [0.3, 0.6, 0.1]^T$ and $\sigma^2 = 3 \times 10^{-3}$ (SNR $= L^{-1}\sigma^{-2} \left\|\mathbf{g}_b(\boldsymbol{a})\right\|^2 \simeq 15$dB). Four different values of b have been assumed under hypothesis H_1, i.e., $b = 5\sigma^2$, $b = 10\sigma^2$, $b = 15\sigma^2$ and $b = 20\sigma^2$. The theoretical ROCs are compared with the empirical and approximated ROCs, where the probabilities of detection are defined as

$$P_{\text{D}}^{\text{emp}}(\eta) = \frac{1}{N}\sum_{n=1}^{N} d_n(\eta) \tag{3.28}$$

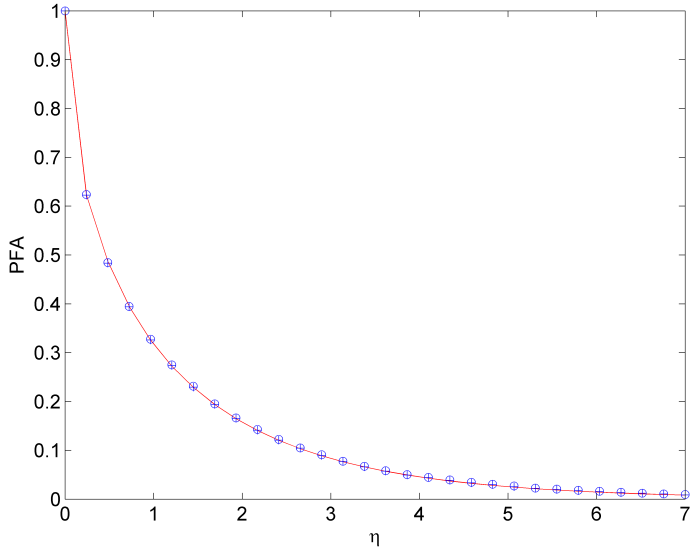


Figure 3.5: Actual (red lines), empirical (blue plus) and approximated (blue circles) PFAs.

and

$$P_{\text{D}}^{\text{app}}(\eta^*) = \frac{1}{N}\sum_{n=1}^{N} \hat{d}_n(\eta) \tag{3.29}$$

and where the data have been generated according to hypothesis H_1. These results show that the proposed test provides similar performance when compared to the original likelihood ratio test (assuming the actual parameters $\boldsymbol{a}$, b and σ^2 are known).

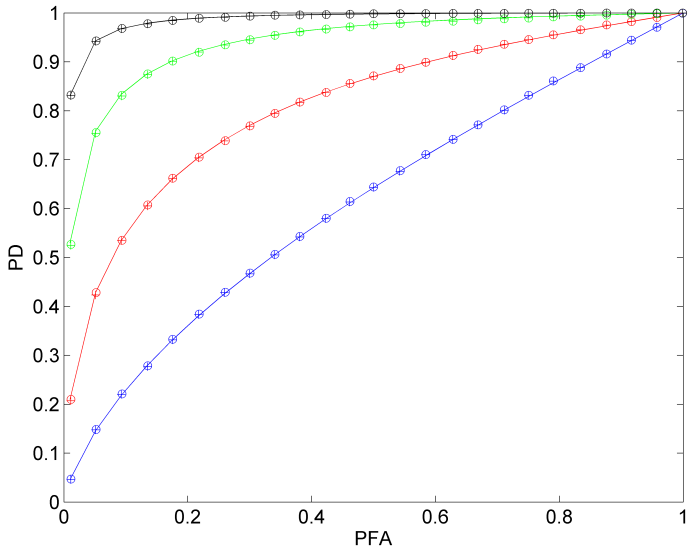


Figure 3.6: Actual (lines), empirical (plus) and approximated (circles) receiver operating characteristics (ROCs) for $b = 5\sigma^2$ (blue), $b = 10\sigma^2$ (red), $b = 15\sigma^2$ (green) and $b = 20\sigma^2$(black).

The performance of the proposed nonlinearity detector is also investigated by testing independently each pixel of a 100×100 synthetic image generated according to the PPNMM. The abundance vectors $\boldsymbol{a}_n, n = 1, \ldots, 10000$, have been randomly drawn from a uniform distribution in the simplex defined by the positivity and sum-to-one constraints. All pixels have been corrupted by an additive Gaussian noise of variance $\sigma^2 = 3 \times 10^{-3}$, corresponding to SNR $\simeq$ 15dB. The nonlinearity parameters have been chosen in the set $\{5\sigma^2, 10\sigma^2, 20\sigma^2, 30\sigma^2\}$, defining four different nonlinearity levels. Fig. 3.7 presents the actual nonlinearity parameters and the detection maps using the subgradient-based estimation procedure for $P_{\text{FA}} = 0.01$ and $P_{\text{FA}} = 0.05$. The white (resp. black) pixels are detected as nonlinearly (resp. linearly) distributed pixels. Note that similar results would be obtained when using the Taylor-based estimation procedure.

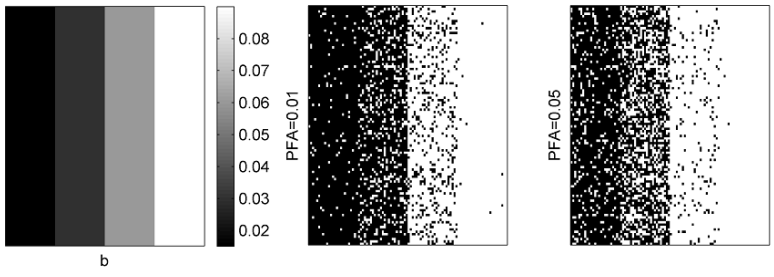


Figure 3.7: Actual values of b (left) and detection maps for $\text{P}_{\text{FA}} = 0.01$ (middle) and $\text{P}_{\text{FA}} = 0.05$ (right) using the subgradient-based algorithm. Black (resp. white) pixels correspond to pixels detected as linearly (resp. nonlinearly) mixed.

The capacity of the PPNMM to detect various nonlinearities is then investigated by unmixing a 100×100 synthetic

image generated according to four different mixing models. The $R = 3$ endmembers contained in this image have been extracted from the spectral libraries provided with the ENVI software (RSI (Research Systems Inc.), 2003) (i.e., green grass, olive green paint and galvanized steel metal). The considered image is divided into four 50×50 sub-images as follows. The first synthetic sub-image $\mathcal{S}_1$ has been generated using the standard linear mixing model (LMM). A second sub-image $\mathcal{S}_2$ has been generated according to the bilinear mixing model introduced in (Fan et al., 2009), referred to as "Fan model" (FM). A third sub-image $\mathcal{S}_3$ has been generated according to the generalized bilinear mixing model (GBM) introduced in (Halimi et al., 2011a,b), whereas a fourth sub-image $\mathcal{S}_4$ has been generated according to the proposed PPNMM. For each sub-image, the abundance vectors $\boldsymbol{a}_n, n = 1, \ldots, 2500$, have been randomly generated according to a uniform distribution in the admissible set defined by the positivity and sum-to-one constraints. All sub-images have been corrupted by an additive white Gaussian noise corresponding to SNR = 15dB. The nonlinearity coefficients are uniformly drawn in the set $(0, 1)$ for the GBM and the parameter b has been generated uniformly in the set $(-0.3, 0.3)$ for the PPNMM.

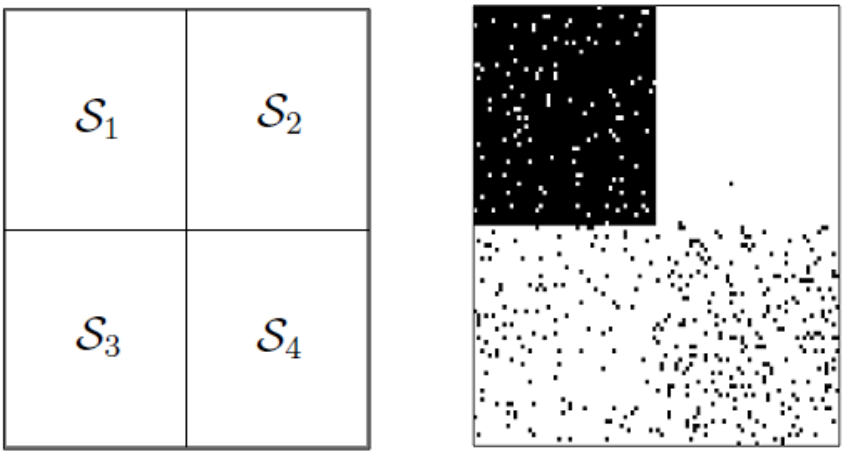


Figure 3.8: Left: Actual location of the four sub-images S_1 (LMM), S_2 (FM), S_3 (GBM) and S_4 (PPNMM). Right: Associated detection map using the subgradient-based algorithm. Black (resp. white) pixels correspond to pixels detected as linearly (resp. nonlinearly) mixed.

Fig. 3.8 (right) shows the detection maps obtained with the GLRT for $P_{\text{FA}} = 0.05$. From this figure, it can be seen that the location of the nonlinear mixtures on the detection maps is straightforward. Note that for the GBM and the PPNMM, mixed pixels can be close to the simplex corresponding to the noise-free LMM and can be detected as linearly distributed pixels. Conversely, for the FM, only almost pure pixels are close to that simplex, leading to a larger number of pixels detected as nonlinear. This remark is illustrated in Fig. 3.9 which shows the location of the pixels detected as nonlinear in the 3-dimensional subspace spanned by the three dominant axes resulting from a principal component analysis.

3.3.5 Analysis of real data

The performance of the proposed nonlinearity detector has been evaluated on a real hyperspectral image composed of $L = 189$ spectral bands. The selected scene has been extracted from the AVIRIS Cuprite image, acquired over

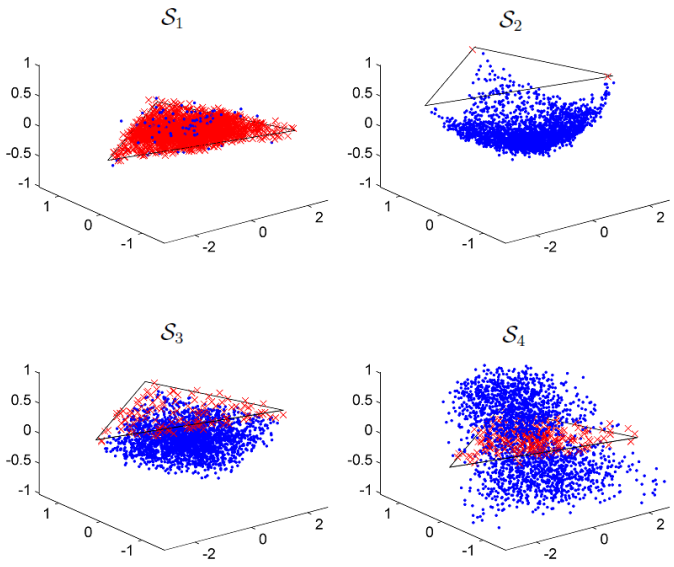


Figure 3.9: Pixels detected as linear (red crosses) and nonlinear (blue dots) for the four sub-images S_1 (LMM), S_2 (FM), S_3 (GBM) and S_4 (PPNMM). The simplex corresponding to the noise-free case LMM is depicted in black lines.

a mining site in Nevada, in 1997. The geologic characteristics of the complete data have been described in (Clark *et al.*, 2003). The area of interest of size 190×250 is represented in Fig. 3.10 and has been previously studied by Nascimento and Bioucas-Dias (2005) to test the VCA algorithm with $R = 14$ endmembers. Therefore, in this experiment, the same number of endmembers has been extracted by VCA. The subgradient-based estimator has

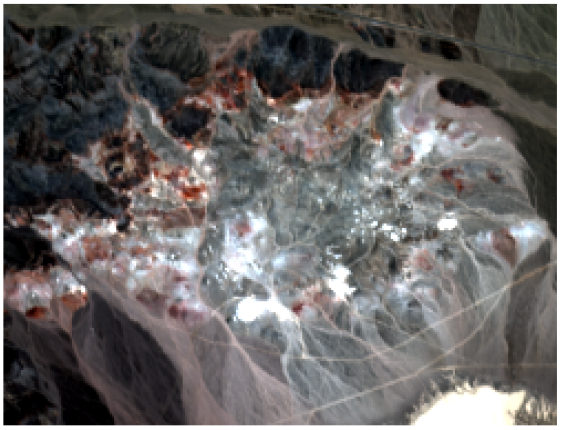


Figure 3.10: AVIRIS image of 190×250 pixels extracted from Cuprite scene observed in composite natural colors.

been used to estimate the parameters of the PPNMM related to the analyzed image, i.e., the abundance vectors, the nonlinearity parameters and the noise variances associated with all image pixels. Fig. 3.11 shows the abundance maps corresponding to the $R = 14$ components. The proportions of pure materials obtained with the PPNMM are in good agreement with those obtained with the FCLS algorithm and depicted in Fig. 3.12.

However, the PPNMM has the advantage of providing additional information regarding the linearity or nonlinearity

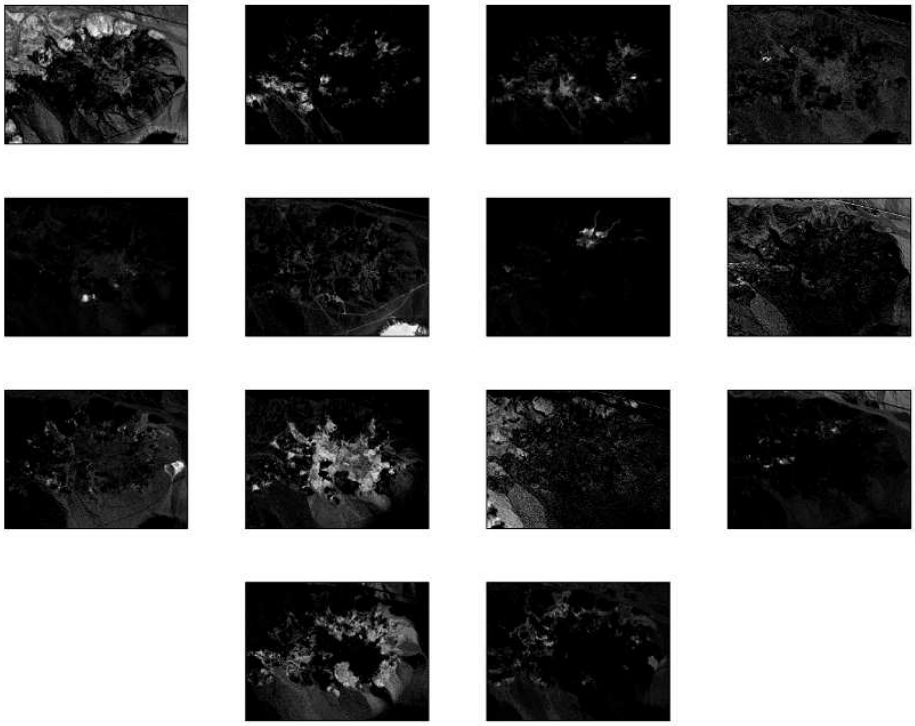


Figure 3.11: Fourteen abundance maps estimated with the subgradient algorithm for the Cuprite scene.

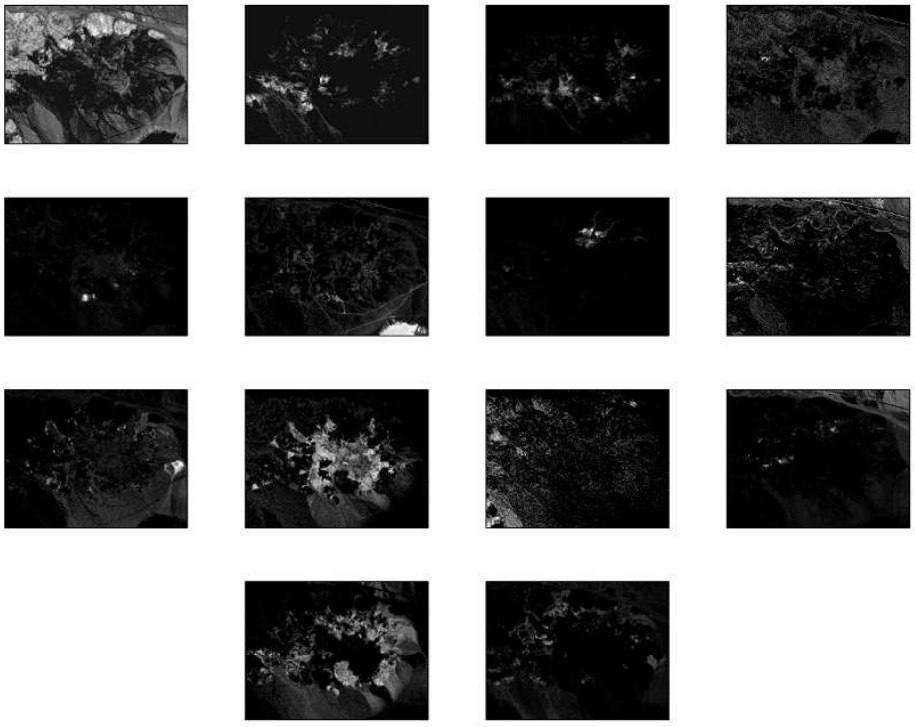


Figure 3.12: Fourteen abundance maps estimated with the FCLS algorithm for the Cuprite scene.

of endmember mixtures via the nonlinearity parameter b. Fig. 3.13 shows the estimated nonlinearity parameter map for the Cuprite scene of Fig. 3.10. Examples of decision maps associated with the subgradient-based estimator

are also depicted in this figure. These decision maps have been obtained by applying the test (3.16) for all pixels of the image for two PFAs. Fig. 3.13 highlights some structures, e.g., the road is clearly identified at the top right corner, especially for $P_{\text{FA}} = 10^{-2}$. A spread nonlinear area is also detected (at the bottom left corner of the image). It can be noted from the classification map of (Clark et al., 1993) that this area is mainly composed of several kinds of Kaolinite. The proposed nonlinearity detector shows that nonlinear effects occur between the different kinds of Kaolinite in this area.

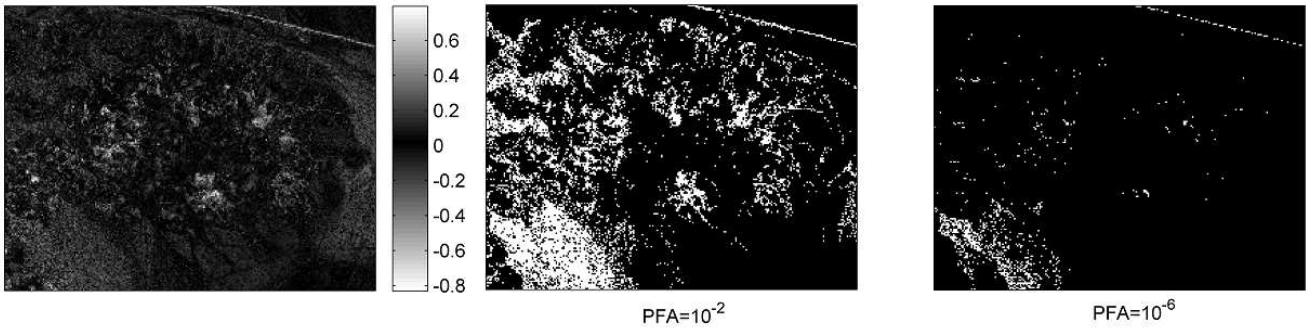


Figure 3.13: Left: map of $\hat{b}$ for the Cuprite scene. Associated detection map for $P_{\text{FA}} = 10^{-2}$ (middle) and for $P_{\text{FA}} = 10^{-6}$ (right). Black (resp. white) pixels correspond to pixels detected as linearly (resp. nonlinearly) mixed.

3.3.6 Intermediate conclusion

A first nonlinearity detector was presented for hyperspectral image analysis. This detector decided if a pixel of a hyperspectral image is a linear combination of endmembers or results from a general nonlinear mixture. It assumed that the hyperspectral image pixels are related to the endmembers by a polynomial post-nonlinear mixing model generalizing the widely used linear mixing model. A subgradient-based algorithm was used to estimate the model parameters. Constrained Cramér-Rao lower bounds were also derived for the PPNMM parameters. These bounds provide a reference in term of estimation variance for estimators satisfying the positivity and sum-to-one constraints of the abundances. The bound for the nonlinearity parameter was also used to approximate the variance of the its maximum likelihood estimator and to build a nonlinearity detector for hyperspectral images. Results obtained on synthetic and real images illustrated the accuracy of the polynomial post-nonlinear model for detecting nonlinearities in hyperspectral images. When a material is too rarely represented or too mixed in the image, it becomes difficult or even impossible to identify its spectrum using an EEA. Simulations conducted on a real image have shown that the proposed nonlinearity detector allows such materials to be identified (such as the road in the Cuprite image).

As mentioned above, the proposed nonlinearity detector relies on an explicit characterization of the nonlinearities, through the consideration of a post-nonlinear model. Even if this algorithm can detect unidentified endmembers, it is not dedicated to this purpose. The next part of this chapter studies a nonlinearity detector for hyperspectral image analysis which does not enforce any particular form of nonlinearities in order to detect more general nonlinearities.

3.4 Supervised LMM-based nonlinearity detection

Similarly to the nonlinearity detector studied in the previous paragraph, this paragraph addresses the problem of determining whether an observed pixel of a hyperspectral image is a linear function of endmembers or results from a generic nonlinear mixing. Differently from paragraph 3.3, the proposed nonlinearity detector is based only on the LMM associated with the null hypothesis. This makes the detector more robust to the nonlinearity that defines the alternative hypothesis. Under the null hypothesis, the linear mixture of the endmembers belongs to a low-dimensional hyperplane. Hence, we propose to design a statistical test based on the distance between the observed pixel and that hyperplane for deciding between the null and alternative hypotheses. As in paragraph 3.3, we assume that the endmembers contained in the image are known.

3.4.1 Mixing models

Let $\mathbf{y}$ be the $L \times 1$ pixel observed in L different spectral bands. The LMM assumes that $\mathbf{y}$ results from a mixture of R known endmembers $\mathbf{m}_1, \ldots, \mathbf{m}_R$ as follows

$$\mathbf{y} = \mathbf{M}\boldsymbol{a} + \mathbf{e} \tag{3.30}$$

where $\mathbf{M} = [\mathbf{m}_1, \ldots, \mathbf{m}_R]$ is the $L \times R$ endmember matrix, the elements a_r of $\boldsymbol{a} = [a_1, \ldots, a_R]^T$ are the proportions of each endmember in the mixture and $\mathbf{e}$ is an $L \times 1$ independent white Gaussian vector such that $\mathbf{e} \sim \mathcal{N}\left(\mathbf{0}_L, \sigma^2 \mathbf{I}_L\right)$. The elements of the abundance vector $\boldsymbol{a}$ satisfy the physical positivity and sum-to-one constraints

$$\sum_{r=1}^{R} a_r = 1, \quad a_r \geq 0, \forall r \in \{1, \ldots, R\}. \tag{3.31}$$

Consider now the hyperplane $\mathcal{H}$ defined by

$$\mathcal{H} : \left\{ \mathbf{z} \middle| \mathbf{z} = \mathbf{M}\boldsymbol{a}, \sum_{r=1}^{R} a_r = 1 \right\}. \tag{3.32}$$

In the noise-free case, $\mathcal{H}$ lies in an $(R-1)$-dimensional subspace embedding all observations distributed according to the LMM.

As in Dobigeon and Févotte (2013), we also consider a general nonlinear mixing model as follows

$$\mathbf{y} = \mathbf{M}\boldsymbol{a} + \boldsymbol{\mu} + \mathbf{e} \tag{3.33}$$

where $\boldsymbol{\mu}$ is an $L \times 1$ deterministic vector that does not belong to $\mathcal{H}$, i.e., $\boldsymbol{\mu} \notin \mathcal{H}$ and $\boldsymbol{a}$ satisfies the constraints (4.2). Note that $\boldsymbol{\mu}$ can be a nonlinear function of the endmember matrix $\mathbf{M}$ and/or the abundance vector $\boldsymbol{a}$ and should be denoted as $\boldsymbol{\mu}(\mathbf{M}, \boldsymbol{a})$. However, the arguments $\mathbf{M}$ and $\boldsymbol{a}$ are omitted in this study for brevity.

Given an observation vector $\mathbf{y}$, we formulate the detection of nonlinear mixtures as the following binary hypothesis testing problem:

$$\begin{cases} H_0 : & \mathbf{y} \text{ is distributed according to (3.30)} \\ H_1 : & \mathbf{y} \text{ is distributed according to (3.33).} \end{cases} \tag{3.34}$$

Using the statistical properties of the noise $\mathbf{e}$, we obtain $\mathrm{E}[\mathbf{y}|H_0] = \mathbf{M}\boldsymbol{a} \in \mathcal{H}$ whereas $\mathrm{E}[\mathbf{y}|H_1] = \mathbf{M}\boldsymbol{a} + \boldsymbol{\mu} \notin \mathcal{H}$. As a consequence, it makes sense to consider the squared Euclidean distance

$$\delta^2(\mathbf{y}) = \min_{\mathbf{z} \in \mathcal{H}} \|\mathbf{y} - \mathbf{z}\|^2 \tag{3.35}$$

between the observed pixel $\mathbf{y}$ and the hyperplane $\mathcal{H}$ to decide which hypothesis (H_0 or H_1) is true. The next paragraph studies the distribution of $\delta^2(\mathbf{y})$ under the two hypotheses H_0 and H_1.

3.4.2 Distributions of $\delta^2(\mathbf{y})$ under hypotheses H_0 and H_1

We now design a statistical test for hypotheses H_0 and H_1 by studying the distribution of $\delta^2(\mathbf{y})$ under each hypothesis. Given the sum-to-one constraint on the abundance vector, the mixing model (3.33) can be rewritten as

$$\tilde{\mathbf{y}} = \mathbf{y} - \mathbf{m}_R = \mathbf{K}\mathbf{c} + \boldsymbol{\mu} + \mathbf{e} \tag{3.36}$$

where $\mathbf{c} = [a_1, \ldots, a_{R-1}]^T$ is $(R-1) \times 1$ and $\mathbf{K} = [\mathbf{m}_1 - \mathbf{m}_R, \ldots, \mathbf{m}_{R-1} - \mathbf{m}_R]$ is $L \times (R-1)$. Hence, $\delta^2(\mathbf{y})$ can be computed by solving the unconstrained least squares (LS) problem

$$\delta^2(\mathbf{y}) = \min_{\mathbf{c} \in \mathbb{R}^{(R-1)}} \|\tilde{\mathbf{y}} - \mathbf{K}\mathbf{c}\|^2 . \tag{3.37}$$

It is well known that the solution to this problem is given by

$$\hat{\mathbf{c}}_{\mathrm{LS}} = \left(\mathbf{K}^T\mathbf{K}\right)^{-1}\mathbf{K}^T\tilde{\mathbf{y}} \tag{3.38}$$

yielding the following $L \times 1$ residual vector $\hat{\mathbf{e}}$

$$\begin{aligned} \hat{\mathbf{e}} &= \tilde{\mathbf{y}} - \mathbf{K}\hat{\mathbf{c}}_{\mathrm{LS}} \\ &= \left[\mathbf{I}_L - \mathbf{K}\left(\mathbf{K}^T\mathbf{K}\right)^{-1}\mathbf{K}^T\right]\tilde{\mathbf{y}} \\ &= \mathbf{H}\left(\boldsymbol{\mu} + \mathbf{e}\right) \end{aligned} \tag{3.39}$$

where $\mathbf{H} = \mathbf{I}_L - \mathbf{K}\left(\mathbf{K}^T\mathbf{K}\right)^{-1}\mathbf{K}^T$ is an $L \times L$ projection matrix of rank $K = L - R + 1$. Using $\delta^2(\mathbf{y}) = \hat{\mathbf{e}}^T\hat{\mathbf{e}}$ and $\hat{\mathbf{e}} \sim \mathcal{N}\left(\mathbf{H}\boldsymbol{\mu}, \sigma^2\mathbf{H}\right)$, straightforward computations lead to (Papoulis, 1991)

$$\frac{1}{\sigma^2}\delta^2(\mathbf{y})|H_1 \sim \chi^2_K\left(\frac{\boldsymbol{\mu}^T\mathbf{H}\boldsymbol{\mu}}{\sigma^2}\right) \tag{3.40}$$

where $\chi^2_K(\lambda)$ denotes the noncentral χ^2 distribution with K degrees of freedom and noncentrality parameter λ. The distribution of $\delta^2(\mathbf{y})$ under H_0 can be obtained by setting $\boldsymbol{\mu} = \mathbf{0}$ in (3.40), yielding

$$\frac{1}{\sigma^2}\delta^2(\mathbf{y})|H_0 \sim \chi^2_K(0) = \chi^2_K \tag{3.41}$$

where χ^2_K is the χ^2 distribution with K degrees of freedom. Notice that the distributions of $\delta^2(\mathbf{y})$ under H_0 and H_1 depend on the known matrix $\mathbf{M}$ but also on the usually unknown noise variance σ^2 and nonlinearity vector $\boldsymbol{\mu}$. In the following we study nonlinearity detectors constructed for known and unknown noise variance σ^2.

3.4.3 Nonlinearity detection

As mentioned above, the distributions of $\delta^2(\mathbf{y})$ under hypotheses H_0 and H_1 depend on σ^2 and on $\boldsymbol{\mu}$, which are usually unknown. We address first the scenario in which the noise variance is known but the nonlinearity $\boldsymbol{\mu}$ (under H_1) is unknown.

Known σ^2, unknown $\boldsymbol{\mu}$

For known σ^2, the distribution of $\delta^2(\mathbf{y})$ is perfectly known under H_0 and partially known under H_1. Thus, we use a statistical test that does not depend on the unknown nonlinearity $\boldsymbol{\mu}$ to decide between H_0 and H_1. Here we propose to use the following statistical test

$$T = \frac{\delta^2(\mathbf{y})}{\sigma^2} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{3.42}$$

where η is a threshold related to the probability of false alarm (PFA) of the test

$$P_{\text{FA}} = \mathbb{P}\left[T > \eta \middle| H_0\right] \tag{3.43}$$

or equivalently,

$$\eta = \text{F}^{-1}_{\chi^2_K}\left(1 - P_{\text{FA}}\right) \tag{3.44}$$

where $\text{F}^{-1}_{\chi^2_K}$ is the inverse cumulative distribution function of the χ^2_K-distribution. For a given $\boldsymbol{\mu}$, the power $P_{\text{D}}(\boldsymbol{\mu})$ of the test is

$$P_{\text{D}}(\boldsymbol{\mu}) = 1 - \text{F}_{\chi^2_K(\lambda)}(\eta) \tag{3.45}$$

where $\text{F}_{\chi^2_K(\lambda)}$ is the cumulative distribution function of the $\chi^2_K(\lambda)$-distribution and $\lambda = \sigma^{-2}\boldsymbol{\mu}^T\mathbf{H}\boldsymbol{\mu}$. Note that the probability of detection (PD) $P_{\text{D}}(\boldsymbol{\mu})$ is an increasing function of λ for a fixed threshold η. This makes sense, as the higher the power $\boldsymbol{\mu}^T\mathbf{H}\boldsymbol{\mu}$ of the nonlinearity orthogonal to $\mathcal{H}$, the better the detection performance. Moreover, the lower the noise variance, the better the nonlinearity detection. Unfortunately, the noise variance is unknown in most practical applications. In these cases, the test (3.42) cannot be used. We study a nonlinearity detector for unknown σ^2 in the next paragraph.

Unknown σ^2, unknown $\boldsymbol{\mu}$

When σ^2 is unknown, a simple solution is to replace the actual noise variance in (3.42) by its estimate $\hat{\sigma}^2$. This yields the following test

$$T^* = \frac{\delta^2(\mathbf{y})}{\hat{\sigma}^2} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{3.46}$$

where η is the threshold defined in (3.44). The PFA and PD of the test (3.46) are then given by

$$\begin{aligned} P^*_{\text{FA}} &= \mathbb{P}\left[T^* > \eta \middle| H_0\right] = \mathbb{P}\left[T > \frac{\hat{\sigma}^2}{\sigma^2}\eta \middle| H_0\right] \\ P^*_{\text{D}}(\boldsymbol{\mu}) &= \mathbb{P}\left[T^* > \eta \middle| H_1\right] = \mathbb{P}\left[T > \frac{\hat{\sigma}^2}{\sigma^2}\eta \middle| H_1\right]. \end{aligned} \tag{3.47}$$

The better the estimation of σ^2, the closer the distributions of T and T^* and thus the closer the performances of the tests (3.42) and (3.46). Unfortunately, considering the maximum likelihood estimator of σ^2 under H_0 ($\hat{\sigma}^2_{\text{MLE}} = \delta^2(\mathbf{y})/L$) lead to a constant test in (3.46) (i.e., $T^* = L$). Here we propose to estimate σ^2 through an eigenanalysis of the sample covariance matrix of a set of N pixels assumed to share the same noise variance. The estimate $\hat{\sigma}^2$ is then determined as the average of the $p \leq L$ smallest eigenvalues of the sample covariance matrix. The accuracy of the estimator will depend on the choice of p, which will be discussed in the simulation paragraph. For a given P_{FA} and associated threshold η given by (3.43), the distribution of T^* is shifted to the left if σ^2 is overestimated, i.e., if $\hat{\sigma}^2 > \sigma^2$. This will lead to a P^*_{FA} of the test (3.46) that is lower than the P_{FA} of the test (3.42). Conversely, if $\hat{\sigma}^2 < \sigma^2$, the distribution of T^* is shifted to the right leading to $P^*_{\text{FA}} \geq P_{\text{FA}}$. Thus, it seems reasonable to overestimate σ^2 to ensure P^*_{FA} is upper bounded by a P_{FA} fixed by the user. This observation will be used in paragraph 3.4.4 to adjust the value of p. It is interesting to mention that, similarly to the detector studied in paragraph 3.3, one might think of assuming the efficiency of the MLE of σ^2 and of approximating its

variance using the corresponding CRLB to derive a nonlinearity detector (similarly to the MLE of b in paragraph 3.3). Straightforward computations would lead to

$$\hat{\sigma}^2_{\text{MLE}} \sim \mathcal{N}\left(\sigma^2, \frac{2\sigma^4}{L}\right) \tag{3.48}$$

However, since σ^2 is unknown, it should be replaced by its estimator in the CRLB expression and in the mean of the asymptotic Gaussian distribution (3.48), leading to a constant (null) test statistic.

3.4.4 Simulations

Synthetic data: known σ^2, unknown $\boldsymbol{\mu}$

We first investigate the performance of the test (3.42), which assumes σ^2 known. We consider a mixture of $R = 3$ materials (green grass, olive green paint and galvanized steel metal) whose spectral signatures $\mathbf{m}_r$ composed of $L = 826$ bands have been extracted from the spectral libraries provided with the ENVI software (RSI (Research Systems Inc.), 2003). The abundance vector is fixed to $\boldsymbol{a} = [0.3, 0.6, 0.1]^T$ and the noise variance to $\sigma^2 = 10^{-3}$. The nonlinearity $\boldsymbol{\mu}$ is set as follows

$$\boldsymbol{\mu} = \nu \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} a_i a_j \mathbf{m}_i \odot \mathbf{m}_j \tag{3.49}$$

where ν is a scaling factor selected from the set $\{0.4, 0.5, 0.6, 0.7\}$. This nonlinearity corresponds to the generalized bilinear model (GBM) studied in (Halimi et al., 2011a) with $\gamma_{i,j} = \nu$ for all (i, j). The specific form in (3.49) was chosen so that the impact of the nonlinearity is governed by a single parameter ν. Fig. 3.14 shows the theoretical and empirical ($N = 20000$ noise realizations) receiver operating characteristics (ROCs) (Kay, 1998, p. 74-75) for the test (3.42). Each value of ν corresponds to a different noncentrality parameter $\lambda = \sigma^{-2}\boldsymbol{\mu}^T\mathbf{H}\boldsymbol{\mu}$ for the noncentral χ^2 distribution, ranging from $\lambda \approx 49$ to $\lambda \approx 150$. These results confirm that the performance of test (3.42) improves for larger values of λ (or ν).

Synthetic data: unknown σ^2, unknown $\boldsymbol{\mu}$

We now study the performance of the nonlinearity detector when σ^2 is replaced by its estimate $\hat{\sigma}^2$. Fig. 3.15 shows the ROCs of test (3.46) for $\lambda = 70$ and for three values of $\hat{\sigma}^2$: $0.95\sigma^2$, σ^2 and $1.05\sigma^2$. Note that all ROCs coincide. However, different estimates $\hat{\sigma}^2$ correspond to different points on the curve for a fixed PFA. For instance, if the PFA is fixed to $P_{FA} = 0.1$ and the noise variance is correctly estimated, the corresponding P_D is around $P_D \approx 0.65$ (see Fig. 3.15 (middle)). If $\hat{\sigma}^2 = 0.95\sigma^2$, the PFA of test (3.46) rises to $P^*_{\text{FA}} \approx 0.41$, leading to $P^*_{\text{D}} \approx 0.92$. Conversely, if $\hat{\sigma}^2 = 1.05\sigma^2$, the PFA of test (3.46) falls to $P^*_{\text{FA}} \approx 0.01$, leading to $P^*_{\text{D}} \approx 0.27$.

To investigate the ability of the proposed nonlinearity detector to detect different types of nonlinearities, we unmixed a synthetic image composed of $N = 10^4$ pixels generated according to four different mixing models. The $R = 3$ endmembers contained in this image are the same as in paragraph 3.4.4. We considered the following nonlinear mixing model

$$\mathbf{y}(n) = \mathbf{M}\boldsymbol{a}(n) + \cos(\theta)\boldsymbol{\mu}_1(n) + \sin(\theta)\boldsymbol{\mu}_2(n) + \mathbf{e}_n \tag{3.50}$$

for $n = 1, \ldots, N$, where $\mathbf{e}_n$ is a Gaussian noise vector such that $\mathbf{e}_n \sim \mathcal{N}\left(\mathbf{0}_L, \sigma^2\mathbf{I}_L\right)$ and

$$\boldsymbol{\mu}_1(n) = b\left[(\mathbf{M}\boldsymbol{a}(n)) \odot (\mathbf{M}\boldsymbol{a}(n))\right] \tag{3.51}$$

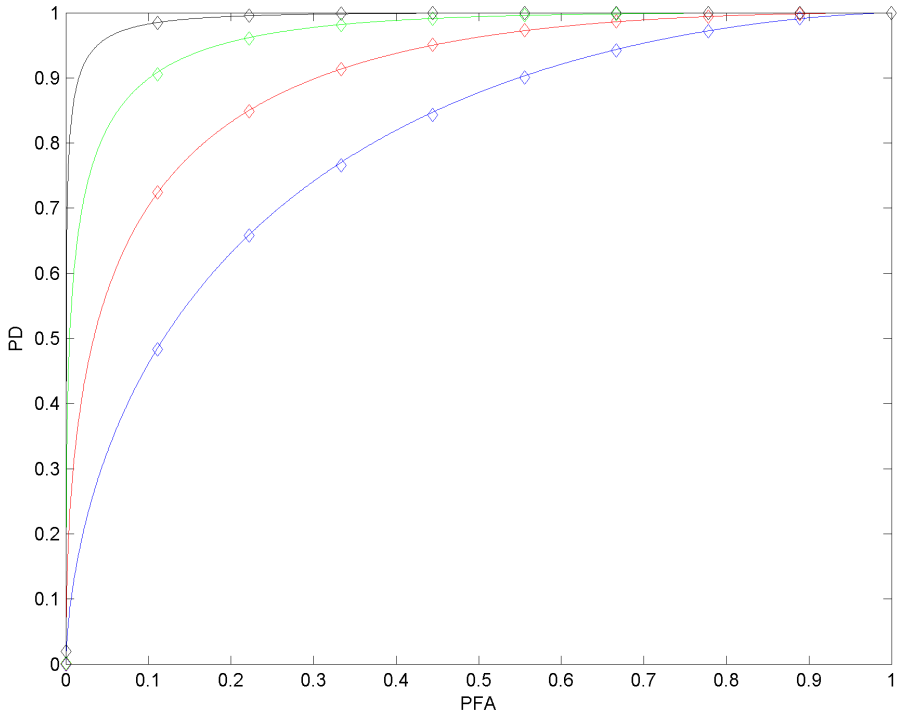


Figure 3.14: Actual (solid lines) and empirical (diamonds) receiver operating characteristics (ROCs) of the first test (known noise variance) for $\nu = 0.4$ (blue), $\nu = 0.5$ (red), $\nu = 0.6$ (green) and $\nu = 0.7$ (black).

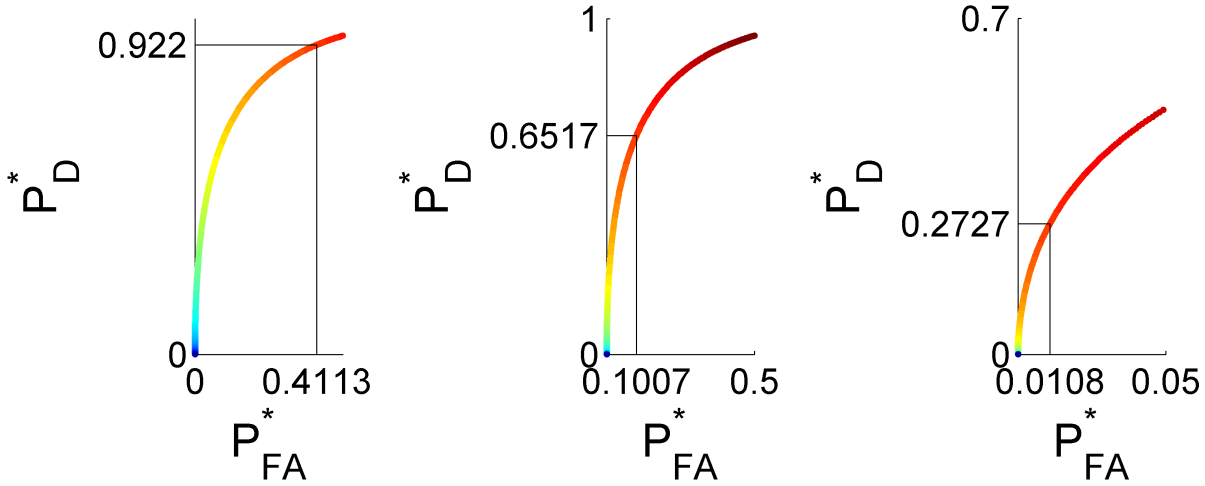


Figure 3.15: Theoretical ROCs of the test (3.46) (unknown noise variance) for $\hat{\sigma}^2 = 0.95\sigma^2$ (left), $\hat{\sigma}^2 = \sigma^2$ (middle) and $\hat{\sigma}^2 = 1.05\sigma^2$ (right). The point corresponding with $P_{FA} = 0.1$ is the intersection of the black lines.

where $\odot$ denotes the Hadamard product and b is a fixed real parameter. The nonlinearity $\boldsymbol{\mu}_1$ corresponds to the nonlinear mixing model studied in paragraph 3.3 for nonlinearity detection. The $L \times 1$ nonlinearity vector

$\boldsymbol{\mu}_2(n)$ is built to ensure that $\boldsymbol{\mu}_2(n)$ is orthogonal to the columns of $\mathbf{M}$ and to $\boldsymbol{\mu}_1(n)$ with $\|\boldsymbol{\mu}_1(n)\| = \|\boldsymbol{\mu}_2(n)\|$ ($\|\cdot\|$ denoting the ℓ_2 norm). The angle $\theta \in [0, \pi/2]$ is chosen to tune the contributions of $\boldsymbol{\mu}_1(n)$ and $\boldsymbol{\mu}_2(n)$ while ensuring $\|\cos(\theta)\boldsymbol{\mu}_1(n) + \sin(\theta)\boldsymbol{\mu}_2(n)\|^2 = \|\boldsymbol{\mu}_1(n)\|^2$ for any value of θ. The main motivation for splitting the nonlinearities into two terms relies on the fact that the PPNMM-based nonlinearity detector cannot identify nonlinearities that are orthogonal to $\boldsymbol{\mu}_1$ since this detector estimates the projection of the each nonlinearity onto $\boldsymbol{\mu}_1$.

The considered set of pixels was divided into four 50×50 sub-images as follows. The first synthetic sub-image $\mathcal{S}_1$ was generated using the standard linear mixing model (LMM) whereas the sub-images $\mathcal{S}_2$, $\mathcal{S}_3$ and $\mathcal{S}_4$ were generated according to the nonlinear mixing model (3.50) with $b = 0.1$ and $\theta \in \{\pi/4, 3\pi/8, \pi/2\}$. For each sub-image, the abundance vectors $\boldsymbol{a}_n, n = 1, \ldots, 2500$, were generated uniformly in the admissible set defined by the positivity and sum-to-one constraints. All sub-images were corrupted by an additive white Gaussian noise of variance $\sigma^2 = 10^{-3}$ corresponding to an average SNR $\approx$ 21dB. Table 3.1 shows the means and standard deviations of the noise variance estimates obtained for different values of p (for 50 Monte Carlo runs). This table shows that fixing $p = L - (R-1)$ provides accurate estimates of σ^2 for these examples. These results can be explained by the fact that the dimension of the subspace spanned by pixels resulting from linear mixtures of R endmembers is $(R-1)$. Fig. 3.16 compares the empirical ROCs constructed from the number of pixels detected as linear and nonlinear for the different tests (known and unknown noise variance). The empirical ROCs for the test studied in paragraph 3.3 are also displayed in these figures. Fig. 3.16 (top left) shows that the three detectors are able to respect the PFA constraint. The three other subfigures of Fig. 3.16 display the ROCs for the three different values of θ. For small values of θ, the norm of the nonlinearity projection onto the vector $\boldsymbol{\mu}_1$ is large. Hence, the nonlinearity detector based on the PPNMM studied in paragraph 3.3 outperforms the tests studied in this paragraph (top right subfigure). However, the performance of test in paragraph 3.3 degrades as the portion of the nonlinearity that is orthogonal to $\boldsymbol{\mu}_1$ becomes predominant (bottom subfigures). The two proposed tests (known and unknown noise variance) perform similarly. Moreover, these two tests seem to be more robust to the type of nonlinearity. Finally, the proposed tests (3.42) and (3.46) only require one projection (3.39) of each pixel (and eventually the noise variance estimation procedure) while the test studied in paragraph 3.3 requires the minimization of a more complex cost function and the derivation of Cramèr-Rao bounds, leading to higher computation costs when compared to the proposed method.

	$\mathcal{S}_1$	$\mathcal{S}_2$	$\mathcal{S}_3$	$\mathcal{S}_4$
$L-p=1$	19.94 (±0.27)	21.32 (±0.35)	20.69 (±0.30)	19.96 (±0.27)
$L-p=2$	**9.99** (±0.01)	**9.98** (±0.01)	**9.99** (±0.01)	**9.99** (±0.01)
$L-p=3$	9.97 (±0.01)	9.97 (±0.01)	9.98 (±0.01)	9.97 (±0.01)
$L-p=4$	9.95 (±0.01)	9.95 (±0.01)	9.96 (±0.01)	9.96 (±0.01)
$L-p=5$	9.94 (±0.01)	9.94 (±0.01)	9.94 (±0.01)	9.94 (±0.01)

Table 3.1: Means and standard deviations (in brackets) of the estimated noise variance ($\times 10^{-4}$) for different values of p ($\sigma^2 = 10^{-3}$).

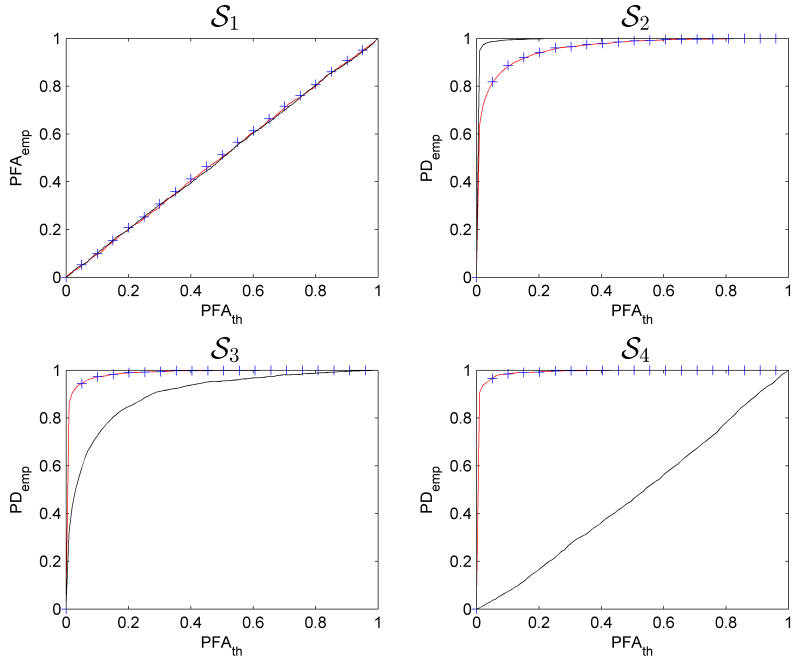


Figure 3.16: Empirical ROCs of the tests (3.42) (red lines), (3.46) (blue crosses) and the test studied in paragraph 3.3 (black lines) for $\mathcal{S}_1$ to $\mathcal{S}_4$.

Real data

The performance of the proposed LMM-based nonlinearity detector has been compared to the PPNMM-based detector studied in paragraph 3.3 using simulations conducted on the 190×250 pixels Cuprite image introduced in paragraph 3.3. The $R = 14$ endmembers have been extracted by VCA. Fig. 3.17 compares the detection maps obtained with the two nonlinearity detectors for $P_{FA} = 10^{-2}$ (left) and $P_{FA} = 10^{-6}$ (right). This figures first shows that the two detectors provide different detection results. The nonlinearly mixed regions identified by the LMM-based detection differs from those obtained with the PPNMM-based detector. Moreover, for a given PFA, the LMM-based detection detects more nonlinearly mixed pixels. Because of the lack of ground truth information about the linear/nonlinear properties of the actual mixtures, it is difficult to decide if one test outperforms the others. However, the two proposed test can be complementary. The LMM-based detector is a general test which does not assume any particular form of nonlinearities while the PPNMM-based detector is more specific and is mainly dedicated to the identification of post-nonlinear mixtures.

3.4.5 Intermediate conclusion

The second nonlinearity detector studied in this chapter was based on the distance between each observed pixel and the low dimensional subspace spanned by the endmembers when the noise variance is known. For an unknown noise variance, a similar detector was proposed by replacing the actual noise variance by an accurate estimator resulting from the eigenanalysis of the sample covariance matrix of a set of image pixels. The main advantages of this method are the absence of prior knowledge about the type of the nonlinearity and its low computational cost. Simulations on synthetic data illustrated the robustness of this method to detect various nonlinearities.

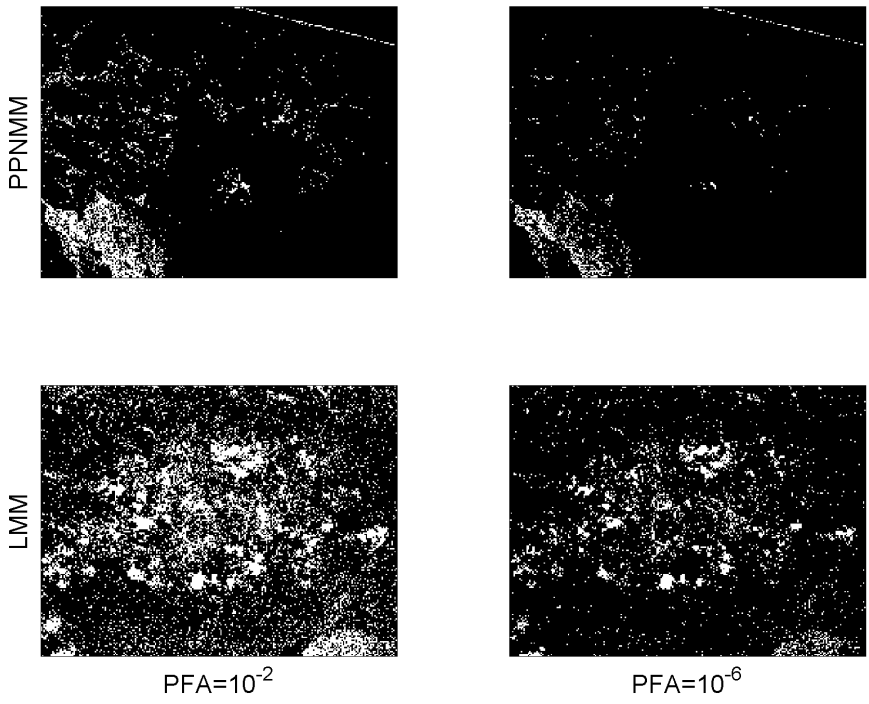


Figure 3.17: Cuprite detection maps obtained with the PPNMM-based (top) and LMM-based (bottom) nonlinearity detectors for $P_{FA} = 10^{-2}$ (left) and $P_{FA} = 10^{-6}$ (right). Black (resp. white) pixels correspond to pixels detected as linearly (resp. nonlinearly) mixed.

3.5 Conclusion

In this chapter, we proposed nonlinearity detectors for applications where the endmembers are known. The first detector was based on the estimated nonlinearity parameter of a polynomial post-nonlinear model. A subgradient-based algorithm was used to estimate the nonlinearity parameter as well as the other model parameters. The variances of these parameter estimators were approximated by the corresponding constrained Cramér-Rao lower bounds allowing to adjust the test threshold as a function of the probability of false alarm, and to compute the probability of detection of the detector.

The second detector was based on the distance between each observed pixel and the subspace spanned by the linear combinations of the endmembers. This detector required to known the noise variance. When this variance was unknown, we proposed to estimate it from the eigenanalysis of the sample covariance matrix of a set of image pixels. The main advantages of the second nonlinearity detector are the absence of prior knowledge about the type of the nonlinearity and its low computational cost. Simulations on synthetic data illustrated the robustness of this method to detect various nonlinearities.

The two proposed detectors have provided promising results in term of detection performance. However, it has been shown that the two detectors can provide different detection results depending on the underlying nonlinearities. It would be interesting to study more complex tests depending on the prior knowledge about the nonlinear effects to be detected. Deviations from the LMM can be explained for instance by colored noise (the noise vectors were assumed to be i.i.d. in this study), endmember estimation errors (when extracted from the data), nonlinear effects

resulting from interactions between the materials of the scene and unidentified endmembers (rarely represented materials). The consideration of such model mis-specifications could provide more specific tests as those developed by Ramsey (1969). To improve the robustness of the detection procedure, several detectors could also be fused.

The two proposed nonlinearity detectors (i.e., the PPNMM-based and the LMM-based) assume the endmembers are known in order to derive a test statistic. However, in practice the spectral signatures are often unknown and must be estimated from the data. When pure pixels are present in the image, EEA can be used before applying the proposed nonlinearity detectors. If the image pixels are too highly mixed (no pure pixel in the image), unsupervised nonlinearity detectors (assuming that the endmembers are unknown) should be used. Precisely, it would interesting to propose joint unmixing and nonlinearity detection algorithms.

Finally, the two proposed nonlinearity detectors were based on pixel-by-pixel spectral analysis. However, it makes sense to consider that nonlinearities occurring in a given pixel can be related to the nonlinearities in neighboring pixels. Consideration of spatial correlation for nonlinearity detection could provide smoother detection results by removing isolated pixels detected as linearly (resp. nonlinearly) mixed, surrounded by pixels detected as nonlinearly (resp. linearly) mixed. The next Chapter introduces a new nonlinear mixing model for joint supervised unmixing and nonlinearity detection based on spatial structures consideration.

Main contributions. Statistical tests were derived for pixel-by-pixel nonlinearity detection when the endmembers are known. The proposed tests were computationally efficient and thus can be implemented in practical applications. The detection results can be used as a pre-processing step to select pixels or regions where more complex models should be used instead of the classical LMM for SU.

3.6 Conclusion (in French)

Dans ce chapitre, nous avons proposé des détecteurs de non-linéarité pixel par pixel pour des applications où les composants sont connus (nombre et spectres). Le premier détecteur est basé sur le paramètre de non-linéarité estimé du modèle PPNMM. Une méthode de gradient a été utilisée pour estimer le paramètre de non-linéarité ainsi que les autres paramètres du modèle (estimation au sens du maximum de vraisemblance). Les variances de ces estimateurs ont été approchées par des bornes de Cramér-Rao contraintes (associées au modèle PPNMM) permettant d'ajuster le seuil du test en fonction de la probabilité de fausse alarme et de calculer la probabilité de détection du détecteur. Le deuxième détecteur est basé sur la distance de chaque pixel observé à l'hyperplan défini par les spectres des composants purs de l'image et la contrainte de somme-à-un des abondances. Ce détecteur suppose que la variance du bruit est connue. Lorsque cette variance est inconnue, nous avons proposé de l'estimer à partir de l' analyse des valeurs propres de la matrice de covariance d'un ensemble de pixels de l'image. Les principaux avantages du second détecteur de non-linéarités sont l'absence de connaissance *a priori* sur le type de non-linéarités à détecter et son faible coût calculatoire. Les simulations sur données synthétiques ont illustré la robustesse de cette méthode pour détecter diverses non-linéarités.

Les deux détecteurs proposés ont donné des résultats prometteurs en termes de performance de détection. Cependant, ils peuvent fournir des résultats de détection différents en fonction des non-linéarités sous-jacentes. Il serait intéressant d'étudier des tests plus complexes en fonction de la connaissance disponible sur les effets non-linéaires à détecter. Les écarts au modèle linéaire classique peuvent être expliquées par exemple par un bruit coloré (les niveaux de bruit dans les différentes bandes spectrales ont été supposés être indépendants et identiquement distribués dans cette étude), les erreurs d'estimation des signatures spectrales (si extraites des données), les effets non-linéaires résultant des interactions entre les matériaux de la scène ou par des composants non-identifiés (matériaux rarement représentés). L'examen de ces types de deviation au modèle linéaire pourrait fournir des tests plus spécifiques, comme ceux développés par Ramsey (1969). Afin d'améliorer la robustesse du procédé de détection, plusieurs détecteurs pourraient également être utilisés en parallèle.

Les deux détecteurs de non-linéarités proposés (basés sur le modèle PPNMM et le modèle linéaire) supposent que les signatures spectrales des composants sont connues pour en déduire une statistique de test. Cependant, dans la pratique, les signatures spectrales sont souvent inconnues et doivent être estimées à partir des données. Lorsque des pixels purs sont présents dans l'image, un algorithme d'extraction de signatures spectrales peut être utilisé avant l'application des détecteurs de non-linéarités proposés. Si les pixels de l'image sont trop fortement mélangés (aucun pixel pur dans l'image), des détecteurs de non-linéarités non-supervisés (qui ne nécessitent pas de connaitre parfaitement les signatures spectrales des composants) doivent être utilisés. En particulier, il serait intéressant de proposer des algorithmes permettant d'estimer les signatures spectrales et d'effectuer la détection de non-linéarités conjointement.

Enfin, les deux détecteurs de non-linéarité proposés sont basés sur une analyse pixel par pixel. Cependant, il est raisonnable de considérer que les non-linéarités qui se produisent dans un pixel donné peuvent être liées aux non-linéarités dans les pixels voisins. Prendre en compte la corrélation spatiale pour la détection de non-linéarités pourrait fournir des résultats de détection plus "lisses" en enlevant les pixels isolés et détectés comme linéairement (resp. non-linéairement) mélangés, entourés par des pixels détectés comme non-linéairement (resp. linéairement) mélangés. Le chapitre suivant présente un nouveau modèle de mélange non-linéaire pour effectuer conjointement la détection de non-linéarités et le démélange supervisé tout en prenant en compte des corrélations spatiales pour

caractériser les non-linéarités.

Contributions majeures. Deux tests statistiques ont été proposés pour la détection non-linéarités pixel par pixel lorsque les signatures spectrales des composants purs présents sont connues. Les tests proposés sont rapides à mettre en place et peuvent donc être utilisés facilement pour des applications pratiques. Les résultats de détection peuvent être utilisés comme une étape de pré-traitement pour sélectionner des pixels ou des régions où des modèles plus complexes doivent être utilisés à la place du modèle de mélange linéaire classique pour résoudre le problème de démélange spectral.

Chapter 4

Joint supervised unmixing and nonlinearity detection using residual component analysis

The first part of this chapter has been adapted from the journal paper (Altmann et al., 2013b) (submitted).

Contents

4.1 Introduction (in French)

Dans le chapitre précédent, un détecteur de non-linéarités pixel par pixel basé sur le modèle PPNMM a été proposé et a donné des résultats intéressants en terme de localisation de régions où des effets non-linéaires apparaissent. Ce détecteur a été construit en utilisant une étape préalable de démélange basée sur le PPNMM. Les propriétés statistiques de l'estimateur du paramètre de non-linéarité de ce modèle ont ensuite été utilisées pour calculer une statistique de test. Dans ce chapitre, nous proposons d'effectuer simultanément le démélange spectral supervisé et la détection de non-linéarités. Nous avons noté dans le chapitre précédent que la prise en compte de structures spatiales dans les images hyperspectrales, déjà considérée par Eches et al. (2011) pour le démélange linéaire, pourrait également être utilisée pour déduire les régions où se produisent des effets non-linéaires. Cette étude présente un nouvel algorithme bayésien supervisé pour effectuer conjointement le démélange (estimation des abondances) et la détection de non-linéarités. Cet algorithme est supervisé dans le sens où les signatures spectrales des composants de l'image sont supposés connues (choisies parmi une bibliothèque spectrale ou extraites à partir des données par un algorithme d'extraction de signatures spectrales). Cet algorithme est basé sur un modèle de mélange non-linéaire inspiré de l'analyse en composantes résiduelles (RCA) (Kalaitzis and Lawrence, 2012). Dans le contexte du démélange d'images hyperspectrales, les effets non-linéaires sont modélisés par des termes de perturbation additifs, caractérisés par des processus gaussiens. Cela permet aux termes non-linéaires d'être marginalisés, ce qui donne un modèle flexible dépendant uniquement des énergies des non-linéarités. Contrairement aux détecteurs de non-linéarités étudiés dans le chapitre 3, les non-linéarités sont supposées aléatoires. L'image hyperspectrale à analyser est partitionnée en régions homogènes dans lesquelles les non-linéarités sont issues d'un même processus gaussien. Cet algorithme repose sur une classification explicite des pixels de l'image, modélisée par des étiquettes dont les dépendances spatiales sont modélisées à l'aide d'un champ aléatoire de Potts-Markov. Si deux classes (ou niveaux) étaient utilisées (mélanges linéaires vs non linéaires), la carte de détection serait binaire. Cependant, cette méthode permet d'identifier différentes régions où des effets non-linéaires apparaissent, à partir des niveaux d'énergie des effets non-linéaires. Plus précisément, l'algorithme proposé permet d'identifier des régions avec différents niveaux de non-linéarités et caractérisées par différents processus gaussiens. Les détecteurs de non-linéarités étudiés dans le chapitre 3 et la plupart des algorithmes de démélange spectral supposent un bruit additif gaussien, indépendamment et identiquement distribués (i.i.d.) spectralement. Cependant, les récents travaux menés sur des images hyperspectrales réelles ont montré que la variance du bruit peut varier significativement d'une bande spectrale à l'autre. Par conséquent, le fait que la puissance du bruit peut changer avec les longueurs d'ondes est pris en compte dans ce chapitre pour améliorer les performances du démélange et de la détection de la non-linéarités.

Dans le cadre bayésien, des lois *a priori* appropriées sont choisies pour les paramètres inconnus du modèle RCA proposé, à savoir, les abondances, les hyperparamètres des processus gaussiens, les étiquettes des différentes classes et la matrice de covariance du bruit. La loi *a posteriori* jointe de ces paramètres est ensuite calculée. Toutefois, les estimateurs bayésiens classiques ne peuvent être facilement calculés à partir de cette loi *a posteriori* (principalement à cause des contraintes sur les paramètres à estimer et la non-linéarité du modèle). Pour résoudre ce problème, une méthode de simulation de type MCMC est utilisée pour générer des échantillons distribués suivant la loi *a posteriori* d'intérêt. Enfin, les échantillons générés sont utilisés pour calculer les estimateurs bayésiens ainsi que des mesures d'incertitudes telles que des intervalles de confiance.

4.2 Introduction

In the previous chapter, a pixel-by-pixel nonlinearity detector based on the PPNMM has been proposed and has provided interesting results. This detector has been constructed using a PPNMM-based SU procedure. The statistical properties of the nonlinearity parameter estimator of this model were used to subsequently derive an accurate test statistic. Conversely, in this chapter we propose to simultaneously achieve the SU and nonlinearity detection. We have noted that the consideration of spatial structures in the image, already used by Eches et al. (2011) for linear SU, could also be used to infer the locations where nonlinear effects occur. This study presents a new supervised Bayesian algorithm for joint nonlinear SU and nonlinearity detection. This algorithm is supervised in the sense that the endmembers contained in the image are assumed to be known (chosen from a spectral library or extracted from the data by an endmember extraction algorithm (EEA)). This algorithm is based on a nonlinear mixing model inspired from residual component analysis (RCA) (Kalaitzis and Lawrence, 2012). In the context of SU of hyperspectral images, the nonlinear effects are modeled by additive perturbation terms characterized by Gaussian processes (GPs). This allows the nonlinear terms to be marginalized, yielding a flexible model depending only on the nonlinearity energies. Contrary to the PPNMM-based and LMM-based nonlinearity detectors introduced in Chapter 3, the nonlinearities are assumed to be random. The hyperspectral image to be analyzed is partitioned into homogeneous regions in which the nonlinearities share the same GP. This algorithm relies on an explicit image classification, modeled by labels whose spatial dependencies are modeled using a Potts-Markov random field. Consideration of two classes (linear vs. nonlinear mixtures) would lead to binary detection maps. However, this study allows for different nonlinearly mixed regions to be also identified, based on the energy of the nonlinear effects. More precisely, the proposed algorithm can identify regions with different levels of nonlinearity and characterized by different GPs. The nonlinearity detectors studied in Chapter 3 and most SU algorithms assume additive, independent and identically distributed (i.i.d.) noise sequences. However, based on previous work conducted on real hyperspectral images, non i.i.d. noise vectors are considered in this study to improve the unmixing and nonlinearity detection performances.

In the Bayesian framework, appropriate prior distributions are chosen for the unknown parameters of the proposed RCA model, i.e., the mixing coefficients, the GP hyperparameters, the class labels and the noise covariance matrix. The joint posterior distribution of these parameters is then derived. However, the classical Bayesian estimators cannot be easily computed from this joint posterior. To alleviate this problem, a Markov chain Monte Carlo (MCMC) method is used to generate samples according to the posterior of interest. Finally, the generated samples are used to compute Bayesian estimators as well as measures of uncertainties such as confidence intervals.

4.3 Problem formulation

We consider a set of N observed pixel spectra $\mathbf{y}_n = [y_{n,1}, \ldots, y_{n,L}]^T, n \in \{1, \ldots, N\}$ where L is the number of spectral bands. Each of these spectra is defined as a linear combination of R known endmembers $\mathbf{m}_r$, contaminated by an additional spectrum $\boldsymbol{\phi}_n$ and additive noise

$$\begin{aligned} \mathbf{y}_n &= \sum_{r=1}^{R} a_{r,n}\mathbf{m}_r + \boldsymbol{\phi}_n + \mathbf{e}_n \\ &= \mathbf{M}\boldsymbol{a}_n + \boldsymbol{\phi}_n + \mathbf{e}_n, \quad n = 1, \ldots, N \end{aligned} \tag{4.1}$$

where $\mathbf{m}_r = [m_{r,1}, \ldots, m_{r,L}]^T$ is the spectrum of the rth material present in the scene, $a_{r,n}$ is its corresponding proportion in the nth pixel and $\mathbf{e}_n$ is an additive independently and non identically distributed zero-mean Gaussian noise sequence with diagonal covariance matrix $\mathbf{\Sigma}_0 = \text{diag}\left(\boldsymbol{\sigma}^2\right)$, denoted as $\mathbf{e}_n \sim \mathcal{N}\left(\mathbf{0}_L, \mathbf{\Sigma}_0\right)$, where $\boldsymbol{\sigma}^2 = [\sigma_1^2, \ldots, \sigma_L^2]^T$ is the vector of the L noise variances and $\text{diag}\left(\boldsymbol{\sigma}^2\right)$ is an $L \times L$ diagonal matrix containing the elements of the vector $\boldsymbol{\sigma}^2$. Moreover, the term $\boldsymbol{\phi}_n = [\phi_{1,n}, \ldots, \phi_{L,n}]^T$ in (4.1) is an unknown $L \times 1$ additive perturbation vector modeling nonlinear effects occurring in the nth pixel. Note that the usual matrix and vector notations $\mathbf{M} = [\mathbf{m}_1, \ldots, \mathbf{m}_R]$ and $\boldsymbol{a}_n = [a_{1,n}, \ldots, a_{R,n}]^T$ have been used in the second row of Eq. (4.1). There are several motivations for considering the mixing model (4.1): 1) this model reduces to the classical linear mixing model (LMM) for $\boldsymbol{\phi}_n = \mathbf{0}_L$, 2) the model (4.1) is general enough to handle different of kinds of nonlinearities such as the bilinear model studied by Fan et al. (2009) (FM), the generalized bilinear model (GBM) (Halimi et al., 2011a), and the PPNMM studied for nonlinear spectral unmixing in Chapter 1 and nonlinearity detection in Chapter 3. These models assume that the mixing model consists of a linear contribution of the endmembers, corrupted by at least one additive term characterizing the nonlinear effects. In the proposed model (4.1), all additive terms are gathered in the vector $\boldsymbol{\phi}_n$. Note that this model is similar to the one introduced by Dobigeon and Févotte (2013) and called robust LMM, and to the nonlinear mixing model proposed by Chen et al. (2013b).
The abundance vectors $\boldsymbol{a}_n$ satisfy the following positivity and sum-to-one constraints

$$\sum_{r=1}^{R} a_{r,n} = 1, \quad a_{r,n} > 0, \forall r \in \{1, \ldots, R\}. \tag{4.2}$$

The problem addressed in this chapter consists of the joint estimation of the abundance vectors and the detection of nonlinearly mixed pixels (characterized by $\boldsymbol{\phi}_n \neq \mathbf{0}_L$). The two next paragraphs present the proposed Bayesian model for joint unmixing and nonlinearity detection.

4.4 Bayesian model

The unknown parameter vector associated with the proposed model (4.1) contains the abundances $\mathbf{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N]$ (satisfying the constraints (4.2)), the nonlinear terms of each pixel $\{\boldsymbol{\phi}_n\}_{n=1,\ldots,N}$, and the noise variance vector $\boldsymbol{\sigma}^2$. This section summarizes the likelihood and the parameter priors associated with the parameters of the linear part of the model, i.e., $\mathbf{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N]$ and $\boldsymbol{\sigma}^2$. The prior model for the nonlinearities $\{\boldsymbol{\phi}_n\}_{n=1,\ldots,N}$ will be introduced in paragraph 4.5.

4.4.1 Likelihood

Equation (4.1) shows that $\mathbf{y}_n|\mathbf{M}, \boldsymbol{a}_n, \boldsymbol{\phi}_n, \boldsymbol{\sigma}^2$ is distributed according to a Gaussian distribution with mean $\mathbf{M}\boldsymbol{a}_n + \boldsymbol{\phi}_n$ and covariance matrix $\mathbf{\Sigma}_0$, denoted as $\mathbf{y}_n|\mathbf{M}, \boldsymbol{a}_n, \boldsymbol{\phi}_n, \boldsymbol{\sigma}^2 \sim \mathcal{N}\left(\mathbf{M}\boldsymbol{a}_n + \boldsymbol{\phi}_n, \mathbf{\Sigma}_0\right)$. Assuming independence between the observed pixels, the joint likelihood of the observation matrix $\mathbf{Y}$ can be expressed as

$$f(\mathbf{Y}|\mathbf{M}, \mathbf{A}, \mathbf{\Phi}, \boldsymbol{\sigma}^2) \propto \quad |\mathbf{\Sigma}_0|^{-N/2} \text{etr}\left[-\frac{(\mathbf{Y}-\mathbf{X})^T \mathbf{\Sigma}_0^{-1} (\mathbf{Y}-\mathbf{X})}{2}\right] \tag{4.3}$$

where $\mathbf{\Phi} = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_N]^T$ is an $L \times N$ nonlinearity matrix, $\text{etr}(\cdot)$ denotes the exponential trace and $\mathbf{X} = \mathbf{M}\mathbf{A} + \mathbf{\Phi}$ is an $L \times N$ matrix.

4.4.2 Prior for the abundance matrix A

Using the LMM constraints (4.2), each abundance vector can be written as $\boldsymbol{a}_n = [\mathbf{c}_n^T, a_{R,n}]^T$ with $\mathbf{c}_n = [a_{1,n}, \ldots, a_{R-1,n}]^T$ and $a_{R,n} = 1 - \sum_{r=1}^{R-1} a_{r,n}$. Moreover, these constraints impose that $\mathbf{c}_n$ belongs to the simplex

$$\mathcal{S} = \left\{ \mathbf{c} \left| c_r > 0, \forall r \in 1, \ldots, R-1, \sum_{r=1}^{R-1} c_r < 1 \right. \right\} \tag{4.4}$$

To reflect the lack of prior knowledge about the abundances, we propose to assign noninformative prior distributions to the N vectors $\mathbf{c}_n$. More precisely, the following uniform prior

$$f(\mathbf{c}_n) \propto \mathbf{1}_{\mathcal{S}}(\mathbf{c}_n), \quad n \in \{1, \ldots, N\} \tag{4.5}$$

is assigned to each vector $\mathbf{c}_n$. Assuming prior independence between the N abundance vectors $\{\boldsymbol{a}_n\}_{n=1,\ldots,N}$ leads to the following joint prior distribution

$$f(\mathbf{C}) = \prod_{n=1}^{N} f(\mathbf{c}_n) \tag{4.6}$$

where $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_N]$ is an $(R-1) \times N$ matrix.

4.4.3 Prior for the noise variances

A Jeffreys' prior is chosen for the noise variance of each spectral band σ_ℓ^2

$$f(\sigma_\ell^2) \propto \frac{1}{\sigma_\ell^2} \mathbf{1}_{\mathbb{R}^+}(\sigma_\ell^2) \tag{4.7}$$

which reflects the absence of knowledge for this parameter (see (Bernardo and Smith, 1994) for motivation). Assuming prior independence between the noise variances, we obtain

$$f(\boldsymbol{\sigma}^2) = \prod_{\ell=1}^{L} f(\sigma_\ell^2). \tag{4.8}$$

4.5 Modeling the nonlinearities

We propose in this study to exploit spatial correlations between the pixels of the hyperspectral image to be analyzed. It seems reasonable to assume that nonlinear effects occurring in a given pixel are related to the nonlinear effects present in neighboring pixels. Formally, the hyperspectral image is assumed to be partitioned into K classes denoted as $\mathcal{C}_0, \ldots, \mathcal{C}_{K-1}$. Let $\mathcal{I}_k \subset 1, \ldots, N$ denote the subset of pixel indexes belonging to the kth class ($k = 0, \ldots, K-1$). An $N \times 1$ label vector $\mathbf{z} = [z_1, \ldots, z_N]^T$ with $z_n \in \{0, \ldots, K-1\}$ is introduced to identify the class of each image pixel, i.e.,

$$\mathbf{y}_n \in C_k \Leftrightarrow n \in \mathcal{I}_k \Leftrightarrow z_n = k. \tag{4.9}$$

In each class, nonlinearity vectors to be estimated are assumed to share the same statistical properties, as will be shown in the sequel.

4.5.1 Prior distribution for the nonlinearity matrix $\mathbf{\Phi}$

As mentioned above, the mixing model (4.1) reduces to the LMM for $\boldsymbol{\phi}_n = \mathbf{0}_L$. For nonlinearity detection, it makes sense to consider a pixel class (referred to as class $\mathcal{C}_0$) corresponding to linearly mixed pixels. The resulting prior distribution for $\boldsymbol{\phi}_n$ conditioned upon $z_n = 0$ is given by

$$f(\boldsymbol{\phi}_n | z_n = 0) = \prod_{\ell=1}^{L} \delta(\phi_{\ell,n}). \tag{4.10}$$

It can be seen that bilinear models and more generally polynomial models (i.e., model involving polynomials nonlinearities with respect to the endmembers) are particularly well adapted to model scattering effects, mainly observed in vegetation and urban areas. Consequently, it makes sense to assume that the nonlinearities $\boldsymbol{\phi}_n$ depend on the endmember matrix $\mathbf{M}$. Nonlinear effects can vary, depending on the relief of the scene, the underlying components involved in the mixtures and the observation conditions to name a few factors. This makes the choice of a single informative prior distribution challenging. From a classification point of view, it is interesting to identify regions or classes where similar nonlinearities occur. For these reasons, we propose to divide nonlinearly mixed pixels into $K-1$ classes and to assign different priors for the nonlinearity vectors belonging to the different classes. The nonlinearities (of nonlinearly mixed pixels) are assumed to be random. Assume $\mathbf{y}_n$ belongs to the kth class. The prior distribution of the corresponding nonlinear term $\boldsymbol{\phi}_n$ is given by the following GP ($k = 1, \ldots, K-1$)

$$\boldsymbol{\phi}_n | \mathbf{M}, z_n = k, s_k^2 \sim \mathcal{N}\left(\mathbf{0}_L, s_k^2 \mathbf{K}_{\mathbf{M}}\right), \tag{4.11}$$

where $\mathbf{K}_{\mathbf{M}}$ is an $L \times L$ covariance matrix parameterized by the endmember matrix $\mathbf{M}$ and s_k^2 is a scaling hyperparameter that tunes the energy of the nonlinearities in the kth class. Note that all nonlinearity vectors within the same class share the same prior. The performance of the unmixing procedure depends on the choice of $\mathbf{K}_{\mathbf{M}}$, more precisely on the similarity measure associated with the covariance matrix. In this work, we consider the symmetric second order polynomial kernel, which has received considerable interest in the machine learning community Scholkopf and Smola (2001). This kernel is defined as follows

$$[\mathbf{K}_{\mathbf{M}}]_{i,j} = \left(\mathbf{m}_{i,:}^T \mathbf{m}_{j,:}\right)^2, \quad i,j \in \{1, \ldots, L\}, \tag{4.12}$$

where $\odot$ denotes the Hadamard (termwise) product and $\mathbf{m}_{i,:}$ denotes the ith row of $\mathbf{M}$. Polynomial kernels are particularly well adapted to characterize multiple scattering effects (modeled by polynomial functions of the endmembers). Note that the parametrization of the matrix $\mathbf{K}_{\mathbf{M}}$ in (4.12) only involves bilinear and quadratic terms with respect to the endmembers $\mathbf{m}_r, r = 1, \ldots, R$. More, precisely, the matrix $\mathbf{K}_{\mathbf{M}}$ can be rewritten as

$$\mathbf{K}_{\mathbf{M}} = \mathbf{Q}\mathbf{Q}^T$$

where $\mathbf{Q} = [\mathbf{m}_1 \odot \mathbf{m}_1, \ldots, \mathbf{m}_R \odot \mathbf{m}_R, \sqrt{2}\mathbf{m}_1 \odot \mathbf{m}_2, \ldots, \sqrt{2}\mathbf{m}_{R-1} \odot \mathbf{m}_R]$ is an $L \times R(R+1)/2$ matrix. It can be shown that (4.11) and (4.12) can be obtained by defining $\boldsymbol{\phi}_n$ as a linear combination of terms $\mathbf{m}_i \odot \mathbf{m}_j$ (as in (Halimi et al., 2011a)) and by marginalizing the corresponding coefficients using a Gaussian prior parameterized by s_k^2. Marginalizing these coefficients allows the number of unknown parameters to be significantly reduced, leading to the nonlinearities being characterized by a single parameter s_k^2 (see Appendix H for details). Note also that a polynomial kernel similar to (4.12) has been recently considered by Chen et al. (2013b) and that other kernels such as the Gaussian kernel could be investigated to model other nonlinearities (Kalaitzis and Lawrence, 2012).

4.5.2 Prior distribution for the label vector z

In the context of hyperspectral image analysis, the labels $z_1, \ldots, z_N$ indicate the pixel classes and take values in $\{0, \ldots, K-1\}$ where K is the number of classes and the set $\{z_n\}_{n=1,\ldots,N}$ forms a random field. To exploit the correlation between pixels, a Markov random field is introduced as a prior distribution for z_n given its neighbors $\mathbf{z}_{\mathcal{V}(n)}$, i.e.,

$$f(z_n|\mathbf{z}_{\backslash n}) = f(z_n|\mathbf{z}_{\mathcal{V}(n)}) \tag{4.13}$$

where $\mathcal{V}(n)$ is the neighborhood of the nth pixel and $\mathbf{z}_{\backslash n} = \{z_{n'}\}_{n' \neq n}$. More precisely, this study focuses on the Potts-Markov model since it is very appropriate for hyperspectral image segmentation (Eches et al., 2011). Given a discrete random field $\mathbf{z}$ attached to an image with N pixels, the Hammersley-Clifford theorem yields

$$f(\mathbf{z}) = \frac{1}{G(\beta)} \exp\left[\beta \sum_{n=1}^{N} \sum_{n' \in \mathcal{V}(n)} \delta(z_n - z_{n'})\right] \tag{4.14}$$

where $\beta > 0$ is the granularity coefficient, $G(\beta)$ is a normalizing (or partition) constant and $\delta(\cdot)$ is the Dirac delta function. Several neighborhood structures can be employed to define $\mathcal{V}(n)$. Fig. 4.1 shows two examples of neighborhood structures. The eight pixel structure (or 2-order neighborhood) will be considered in the rest of the chapter.

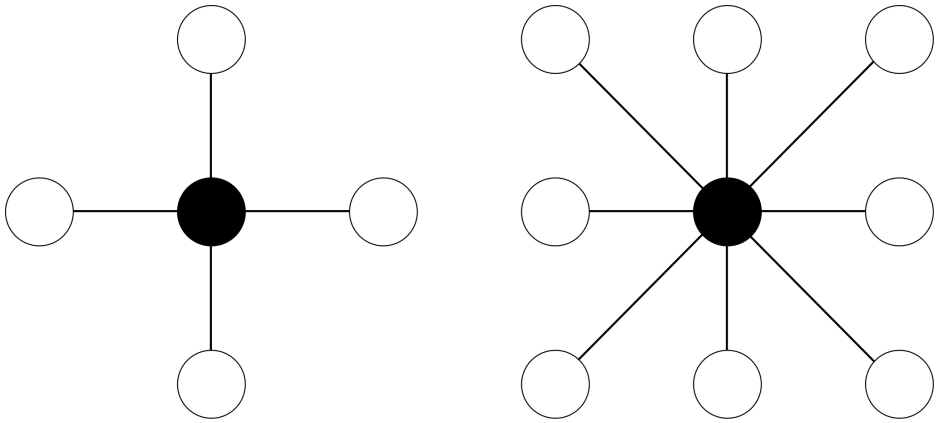


Figure 4.1: 4-pixel (left) and 8-pixel (right) neighborhood structures. The considered pixel appear as a black circle whereas its neighbors are depicted in white.

The hyperparameter β tunes the degree of homogeneity of each region in the image. More precisely, small values of β yield an image with a large number of regions, whereas large values of β lead to fewer and larger homogeneous regions. In this study, the granularity coefficient is assumed to be known. Note however that it could be also included within the Bayesian model and estimated using the strategy described by Pereyra et al. (2013).

4.5.3 Hyperparameter priors

The performance of the proposed Bayesian model for spectral unmixing mainly depends on the values of the hyperparameters $\left\{s_k^2\right\}_{k=1,\ldots,K}$. When the hyperparameters are difficult to adjust, it is the norm to include them in the unknown parameter vector, resulting in a hierarchical Bayesian model (Robert, 2007). This strategy requires the definition of prior distributions for the hyperparameters.

The following inverse-gamma prior distribution

$$s_k^2|\gamma,\nu \sim \mathcal{IG}(\gamma,\nu), \quad \forall k \in \{1,\ldots,K\} \tag{4.15}$$

is assigned to the nonlinearity hyperparameters, where (γ,ν) are additional parameters that will be fixed to ensure a noninformative prior for s_k^2 ($(\gamma,\nu) = (1,1/4)$ in all simulations presented in this chapter). Assuming prior independence between the hyperparameters, we obtain

$$f(\mathbf{s}^2|\gamma,\nu) = \prod_{k=1}^{K-1} f(s_k^2|\gamma,\nu). \tag{4.16}$$

where $\mathbf{s}^2 = [s_1^2,\ldots,s_K^2]^T$.

4.6 Bayesian inference using a Metropolis-within-Gibbs sampler

4.6.1 Marginalized joint posterior distribution

The resulting directed acyclic graph (DAG) associated with the proposed Bayesian model introduced in Sections 4.4 and 4.5 is depicted in Fig. 4.2.

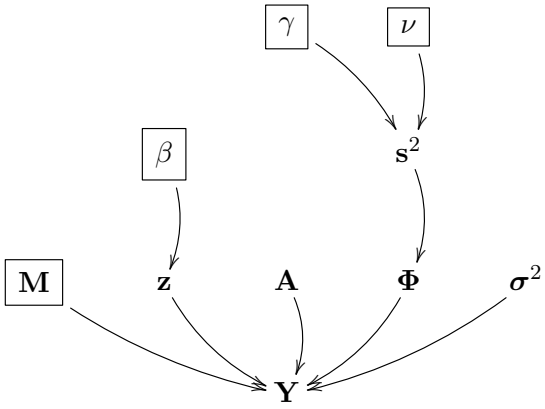


Figure 4.2: DAG for the parameter and hyperparameter priors (the fixed parameters appear in boxes).

Assuming prior independence between $\mathbf{A}$, $(\boldsymbol{\Phi},\mathbf{z})$ and $\boldsymbol{\sigma}^2$, the posterior distribution of $(\boldsymbol{\Phi},\boldsymbol{\theta})$ where $\boldsymbol{\theta} = (\mathbf{C},\mathbf{z},\boldsymbol{\sigma}^2,\mathbf{s}^2)$ can be expressed as

$$f(\boldsymbol{\theta},\boldsymbol{\Phi}|\mathbf{Y},\mathbf{M}) \quad \propto \quad f(\mathbf{Y}|\mathbf{M},\boldsymbol{\theta},\boldsymbol{\Phi}) f(\boldsymbol{\Phi}|\mathbf{M},\mathbf{z},\mathbf{s}^2) f(\boldsymbol{\theta}),$$

where $f(\boldsymbol{\theta}) = f(\mathbf{C}) f(\boldsymbol{\sigma}^2) f(\mathbf{z}) f(\mathbf{s}^2)$. This distribution can be marginalized with respect to $\boldsymbol{\Phi}$ as follows

$$\begin{aligned} f(\boldsymbol{\theta}|\mathbf{Y},\mathbf{M}) &\propto f(\boldsymbol{\theta}) \int f(\mathbf{Y}|\mathbf{M},\boldsymbol{\theta},\boldsymbol{\Phi}) f(\boldsymbol{\Phi}|\mathbf{M},\mathbf{z},\mathbf{s}^2) \mathrm{d}\boldsymbol{\Phi} \\ &\propto f(\boldsymbol{\theta}) f(\mathbf{Y}|\mathbf{M},\boldsymbol{\theta}) \end{aligned} \tag{4.17}$$

where

$$\begin{aligned} f(\mathbf{Y}|\mathbf{M},\boldsymbol{\theta}) &= \int f(\mathbf{Y}|\mathbf{M},\boldsymbol{\theta},\boldsymbol{\Phi}) f(\boldsymbol{\Phi}|\mathbf{M},\mathbf{z},\mathbf{s}^2) \mathrm{d}\boldsymbol{\Phi} \\ &\propto \prod_{k=0}^{K-1} \prod_{n\in\mathcal{I}_k} \frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\bar{\mathbf{y}}_n^T \boldsymbol{\Sigma}_k^{-1} \bar{\mathbf{y}}_n\right] \end{aligned} \tag{4.18}$$

with $\boldsymbol{\Sigma}_0 = \text{diag}\left(\boldsymbol{\sigma}^2\right)$, $\boldsymbol{\Sigma}_k = s_k^2\mathbf{K}_\mathbf{M} + \boldsymbol{\Sigma}_0$ $(k = 1, \ldots, K-1)$ and $\bar{\mathbf{y}}_n = \mathbf{y}_n - \mathbf{M}\boldsymbol{a}_n$. The advantage of this marginalization is to avoid sampling the nonlinearity matrix $\boldsymbol{\Phi}$. Thus, the nonlinearities are fully characterized by the known endmember matrix, the class labels and the values of the hyperparameters in $\mathbf{s}^2 = [s_1^2, \ldots, s_K^2]^T$. Note that the alternative interpretation of the proposed RCA-based model provided in Appendix H also leads to the likelihood marginalized over $\boldsymbol{\Phi}$ in (4.18).

Unfortunately, it is difficult to obtain closed form expressions for standard Bayesian estimators associated with (4.17). In this study, we propose to use efficient Markov Chain Monte Carlo (MCMC) methods to generate samples asymptotically distributed according to (4.17). The next part of this section presents the Gibbs sampler which is proposed to sample according to (4.17). The principle of the Gibbs sampler is to sample according to the conditional distributions of the posterior of interest (Robert and Casella, 2004, Chap. 10). Due to the large number of parameters to be estimated, it makes sense to use a block Gibbs sampler to improve the convergence of the sampling procedure. More precisely, we propose to sample sequentially the N labels in $\mathbf{z}$, the abundance matrix $\mathbf{A}$, the noise variances $\boldsymbol{\sigma}^2$ and $\mathbf{s}^2$ using moves that are detailed in the next paragraphs.

4.6.2 Sampling the labels

For the nth pixel ($n \in \{1, \ldots, N\}$), the label z_n is a discrete random variable whose conditional distribution is fully characterized by the probabilities

$$P(z_n = k|\mathbf{y}_n, \mathbf{M}, \boldsymbol{\theta}_{\backslash z_n}) \quad \propto \quad f(\mathbf{y}_n|\mathbf{M}, \mathbf{s}^2, z_n = k, \boldsymbol{a}_n) f(z_n|\mathbf{z}_{\backslash n}), \tag{4.19}$$

where $\boldsymbol{\theta}_{\backslash z_n}$ denotes $\boldsymbol{\theta}$ without z_n, $k = 0, \ldots, K-1$ (for K classes). These posterior probabilities are

$$P(z_n = k|\mathbf{y}_n, \mathbf{M}, \boldsymbol{\theta}_{\backslash z_n}) \propto \frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \exp\left[\beta \sum_{p=1}^{N} \sum_{p' \in \mathcal{V}(p)} \delta(z_p - z_{p'})\right] \exp\left[-\frac{1}{2}\bar{\mathbf{y}}_n^T \boldsymbol{\Sigma}_k^{-1} \bar{\mathbf{y}}_n\right]. \tag{4.20}$$

Consequently, sampling z_n from its conditional distribution can be achieved by drawing a discrete value in the finite set $\{0, \ldots, K-1\}$ with the probabilities defined in (4.20).

4.6.3 Sampling the abundance matrix A

Sampling from $f(\mathbf{C}|\mathbf{Y}, \mathbf{M}, \mathbf{z}, \boldsymbol{\sigma}^2, \mathbf{s}^2)$ seems difficult due to the complexity of this distribution. However, it can be shown that

$$f(\mathbf{C}|\mathbf{Y}, \mathbf{M}, \mathbf{z}, \boldsymbol{\sigma}^2, \mathbf{s}^2) = \prod_{n=1}^{N} f(\mathbf{c}_n|\mathbf{y}_n, \mathbf{M}, z_n, \boldsymbol{\sigma}^2, \mathbf{s}^2), \tag{4.21}$$

i.e., the N abundance vectors $\{\boldsymbol{a}_n\}_{n=1,\ldots,N}$ are a posteriori independent and can be sampled independently in a parallel manner. Straightforward computations lead to

$$\mathbf{c}_n|\mathbf{y}_n, \mathbf{M}, z_n = k, \boldsymbol{\sigma}^2, \mathbf{s}^2 \sim \mathcal{N}_\mathcal{S}(\bar{\mathbf{c}}_n, \boldsymbol{\Psi}_n) \tag{4.22}$$

where

$$\begin{aligned} \boldsymbol{\Psi}_n &= \left(\widetilde{\mathbf{M}}^T \boldsymbol{\Sigma}_k^{-1} \widetilde{\mathbf{M}}\right)^{-1} \\ \bar{\mathbf{c}}_n &= \boldsymbol{\Psi}_n \widetilde{\mathbf{M}}^T \boldsymbol{\Sigma}_k^{-1} \tilde{\mathbf{y}}_n \\ \widetilde{\mathbf{M}} &= [\mathbf{m}_1 - \mathbf{m}_R, \ldots, \mathbf{m}_{R-1} - \mathbf{m}_R] \end{aligned} \tag{4.23}$$

and $\tilde{\mathbf{y}}_n = \mathbf{y}_n - \mathbf{m}_R$. Moreover, $\mathcal{N}_{\mathcal{S}}(\bar{\mathbf{c}}_n, \boldsymbol{\Psi}_n)$ denotes the truncated multivariate Gaussian distribution defined on the simplex $\mathcal{S}$ with hidden mean $\bar{\mathbf{c}}_n$ and hidden covariance matrix $\boldsymbol{\Psi}_n$. Sampling from (4.22) can be achieved efficiently using the method recently proposed by Pakman and Paninski (2012).

4.6.4 Sampling the noise variance σ^2

It can be shown from (4.17) that

$$f(\boldsymbol{\sigma}^2|\mathbf{Y}, \mathbf{M}, \mathbf{A}, \mathbf{z}, \mathbf{s}^2) = \prod_{\ell=1}^{L} f(\sigma_\ell^2|\mathbf{Y}, \mathbf{M}, \mathbf{A}, \mathbf{z}, \mathbf{s}^2), \tag{4.24}$$

where

$$f(\sigma_\ell^2|\mathbf{Y}, \mathbf{M}, \mathbf{A}, \mathbf{z}, \mathbf{s}^2) \propto \frac{1}{\sigma_\ell^2} \prod_{k=0}^{K-1} \prod_{n \in \mathcal{I}_k} \frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\bar{\mathbf{y}}_n^T \boldsymbol{\Sigma}_k^{-1} \bar{\mathbf{y}}_n\right] \mathbf{1}_{\mathbb{R}^+}\left(\sigma_\ell^2\right). \tag{4.25}$$

Sampling from (4.25) is not straightforward. In this case, an accept/reject procedure can be used to update σ_ℓ^2, leading to a hybrid Metropolis-within-Gibbs sampler. In this study, we introduce the standard change of variables $\delta_\ell = \log(\sigma_\ell^2)$, $\delta_\ell \in \mathbb{R}$. A Gaussian random walk for δ_ℓ is used to update the variance σ_ℓ^2. This change of variables allows the proposals to be symmetric, conversely to the truncated Gaussian distribution. Note that the noise variances σ_ℓ^2 are a posteriori independent. Thus they can be updated in a parallel manner. The variances of the L parallel Gaussian random walk procedures have been adjusted during the burn-in period of the sampler to obtain an acceptance rate close to 0.5, as recommended in (Robert and Cellier, 1998, p. 8).

4.6.5 Sampling the vector $\mathbf{s}^2$

It can be shown from (4.17) that

$$f(\mathbf{s}^2|\mathbf{Y}, \mathbf{M}, \mathbf{A}, \mathbf{z}, \boldsymbol{\sigma}^2, \gamma, \nu) = \prod_{k=1}^{K-1} f(s_k^2|\mathbf{Y}, \mathbf{M}, \mathbf{A}, \boldsymbol{\sigma}^2, \gamma, \nu),$$

where

$$f(s_k^2|\mathbf{Y}, \mathbf{M}, \mathbf{A}, \boldsymbol{\sigma}^2, \gamma, \nu) \propto f(s_k^2|\gamma, \nu) \prod_{n \in \mathcal{I}_k} \frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\bar{\mathbf{y}}_n^T \boldsymbol{\Sigma}_k^{-1} \bar{\mathbf{y}}_n\right]. \tag{4.26}$$

Due to the complexity of the conditional distribution (4.26), Gaussian random walk procedures are used in the log-space to update the hyperparameters $\{s_k^2\}_{k=1,\ldots,K-1}$ in a parallel manner (similarly to the noise variance updates). Again, the proposal variances are adjusted during the burn-in period of the sampler.

After generating N_{MC} samples using the procedures detailed above and removing N_{bi} iterations associated with the burn-in period of the sampler (N_{bi} has been set from preliminary runs), the marginal maximum a posteriori (MAP) estimator of the label vector, denoted as $\hat{\mathbf{z}}_{\text{MAP}}$, can be computed. The label vector estimator is then used to compute the minimum mean square error (MMSE) of $\mathbf{A}$ conditioned upon $\mathbf{z} = \hat{\mathbf{z}}_{\text{MAP}}$. Finally, the noise variances and the hyperparameters $\{s_k^2\}_{k=1,\ldots,K-1}$ are estimated using the empirical averages of the generated samples (MMSE estimates). The next section studies the performance of the proposed algorithm for synthetic hyperspectral images.

4.7 Simulations for synthetic data

4.7.1 First scenario: RCA vs. linear unmixing

The performance of the proposed joint nonlinear SU and nonlinearity detection algorithm is first evaluated by unmixing a synthetic image of 60×60 pixels generated according to the model (4.1). The $R = 3$ endmembers contained in these images (i.e., green grass, olive green paint and galvanized steel metal) have $L = 207$ different spectral bands and have been extracted from the spectral libraries provided with the ENVI software (RSI (Research Systems Inc.), 2003) . The number of classes has been set to $K = 4$, i.e, $K - 1 = 3$ classes of nonlinearly mixed pixels. The hyperparameters $\{s_k^2\}_{k=1,\ldots,3}$ have been fixed as shown in Table 4.2, which represents three possible levels of nonlinearity. For each class, the nonlinear terms have been generated according to (4.11). The label map generated with $\beta = 1.2$ is shown in Fig. 4.3 (left). The abundance vectors $\boldsymbol{a}_n, n = 1, \ldots, 3600$ have been randomly generated according to a uniform distribution over the admissible set defined by the positivity and sum-to-one constraints. The noise variance (depicted in Fig. 4.4 as a function of the spectral bands) have been arbitrarily fixed using

$$\sigma_\ell^2 = 10^{-4}\left[2 - \sin\left(\pi \frac{\ell}{L-1}\right)\right]. \tag{4.27}$$

to model a non-i.i.d. (colored) noise. The joint nonlinear SU and nonlinearity detection algorithm, denoted as "RCA-SU", has been applied to this data set with $N_{\text{MC}} = 3000$ and $N_{\text{bi}} = 1000$. Fig. 4.3 (right) shows that the estimated label map (marginal MAP estimates) is in agreement with the actual label map. Moreover, the confusion matrix depicted in Table 4.1 illustrate the performance of the RCA-SU in term of pixel classification. Table 4.2 shows that the RCA-SU provides accurate hyperparameter estimates and thus can be used to obtain information about the importance of nonlinearities in the different regions. Note that the estimation error is computed using $|s_k^2 - \hat{s}_k^2|/s_k^2$, where s_k^2 and $\hat{s}_k^2$ are the actual and estimated dispersion parameters for the kth class. The estimated noise variances, depicted in Fig. 4.4 are also in good agreement with the actual values of the variances. The

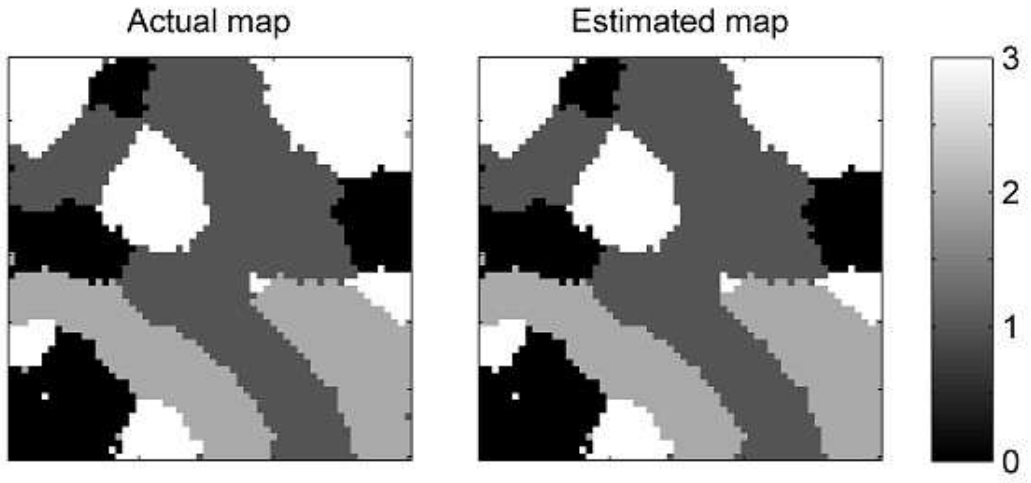


Figure 4.3: Actual (left) and estimated (right) classification maps of the synthetic image associated with the first scenario.

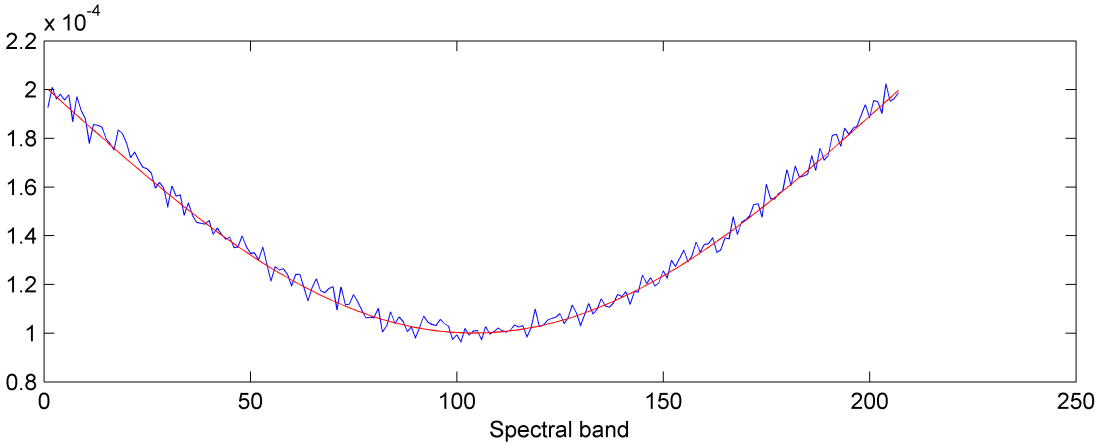


Figure 4.4: Actual noise variances (red) and variances estimated by the RCA-SU algorithm (blue) for the synthetic image associated with the first scenario.

quality of abundance estimation can be evaluated by comparing the estimated and actual abundance vectors using the root normalized mean square error (RNMSE) defined in each class by

$$\text{RNMSE}_k = \sqrt{\frac{1}{N_k R} \sum_{n \in \mathcal{I}_k} \|\hat{\boldsymbol{a}}_n - \boldsymbol{a}_n\|^2} \tag{4.28}$$

with $N_k = \text{card}(\mathcal{I}_k)$ and where $\boldsymbol{a}_n$ and $\hat{\boldsymbol{a}}_n$ are the actual and estimated abundance vectors for the nth pixel of the image. For this scenario, the proposed algorithm is compared with the classical FCLS algorithm (Heinz and C.-I Chang, 2001) assuming the LMM. Comparisons to nonlinear SU methods will be addressed in the next paragraph (scenario 2). Table 4.3 shows the RNMSEs obtained with the proposed and the FLCS algorithms for this first data set. These results show that the two algorithms provide similar abundance estimates for the first class, corresponding to linearly mixed pixels. For the three nonlinear classes, the proposed algorithm provides better results than the FCLS algorithm that does not handle nonlinear effects.

Table 4.1: First scenario: Confusion matrix ($N = 3600$ pixels).

		Estimated classes			
		$\mathcal{C}_0$	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_3$
Actual classes	$\mathcal{C}_0$	659	0	0	0
	$\mathcal{C}_1$	1	1274	2	0
	$\mathcal{C}_2$	0	4	787	2
	$\mathcal{C}_3$	0	0	0	871

Table 4.2: First scenario: Hyperparameter estimation.

	s_1^2	s_2^2	s_3^2
Actual value	0.01	0.1	1
Estimation error	2.76%	1.12%	0.28%

Table 4.3: RNMSEs ($\times 10^{-2}$): synthetic images .

	Class #0	Class #1	Class #2	Class #3
FCLS	0.38	15.23	29.95	42.79
RCA-SU	**0.38**	**2.83**	**3.99**	**4.23**

4.7.2 Second scenario: RCA vs. nonlinear unmixing

Data set

The performance of the proposed joint nonlinear SU and nonlinearity detection algorithm is then evaluated on a second synthetic image of 60×60 pixels containing the $R = 3$ spectral components presented in the previous section. In this scenario, the image consists of pixels generated according to four different mixing models associated with four classes ($K = 4$). The label map generated using $\beta = 1.2$ is shown in Fig. 4.5 (a). The class $\mathcal{C}_0$ is associated with the LMM. The pixels of class $\mathcal{C}_1$ have been generated according to the GBM (Halimi et al., 2011a)

$$\mathbf{y}_n = \sum_{r=1}^{R} a_{r,n}\mathbf{m}_r + \sum_{i=1}^{R-1}\sum_{j=i+1}^{R} \gamma_{i,j} a_{i,n} a_{j,n} \mathbf{m}_i \odot \mathbf{m}_j + \mathbf{e}_n \tag{4.29}$$

where $n \in \mathcal{I}_1$ and the nonlinearity parameters $\{\gamma_{i,j}\}$ have been uniformly drawn in $[0.5, 1]$. The class $\mathcal{C}_2$ is composed of pixels generated according to the PPNMM introduced in Chapter 1 as follows

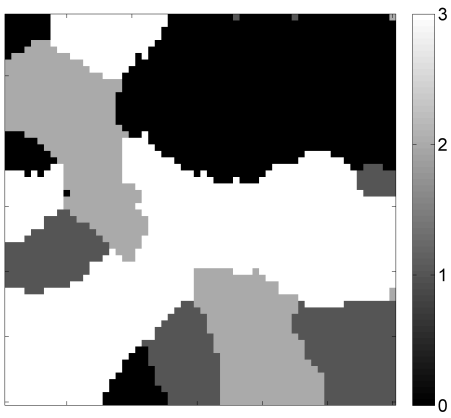
$$\mathbf{y}_n = \sum_{r=1}^{R} a_{r,n}\mathbf{m}_r + b\left(\sum_{r=1}^{R} a_{r,n}\mathbf{m}_r\right) \odot \left(\sum_{r=1}^{R} a_{r,n}\mathbf{m}_r\right) + \mathbf{e}_n \tag{4.30}$$

where $n \in \mathcal{I}_2$ and $b = 0.5$ for all pixels in class $\mathcal{C}_2$. Finally, the class $\mathcal{C}_3$ has been generated according to (4.1) with $s^2 = 0.1$. For the four classes, the abundance vectors have been randomly generated according to a uniform distribution over the admissible set defined by the positivity and sum-to-one constraints. All pixels have been corrupted by an additive i.i.d Gaussian noise of variance $\sigma^2 = 10^{-4}$, corresponding to an average signal-to-noise ratio SNR $\simeq$ 30dB. The noise is assumed to be i.i.d. for a fair comparison with SU algorithms assuming i.i.d. Gaussian noise. Fig. 4.5 (b) shows the log-energy of the nonlinearity parameters for each pixel of the image, i.e., $\log\left(\|\boldsymbol{\phi}_n\|^2\right)$ for $n = 1, \ldots, 3600$. This figure shows that each class corresponds to a different level of nonlinearity.

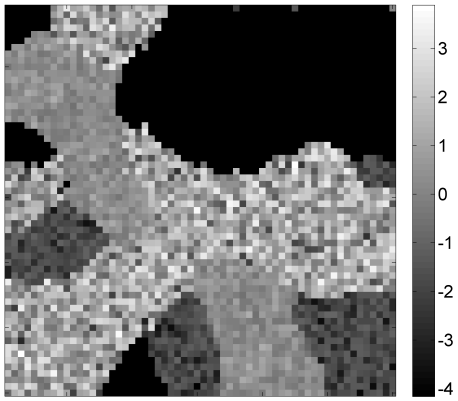
Unmixing

Different estimation procedures have been considered for the four different mixing models:

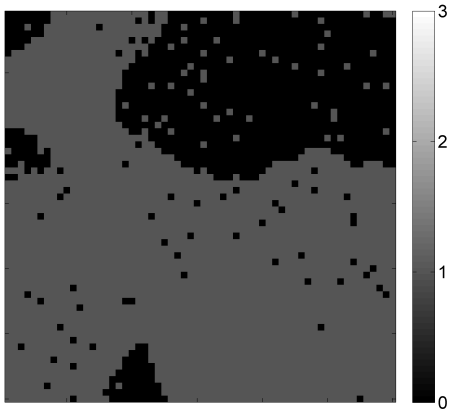
- The FCLS algorithm (Heinz and C.-I Chang, 2001) which is known to have good performance for linear mixtures.
- The GBM-based approach (Halimi et al., 2011b) which is particularly adapted for bilinear nonlinearities.



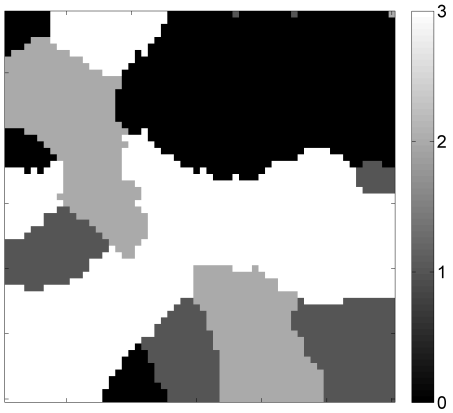
(a) Actual label map.



(b) $\log\left(\|\boldsymbol{\phi}_n\|^2\right)$.



(c) Detection map (PPNMM).



(d) Detection map (RCA-SU).

Figure 4.5: Nonlinearity detection for the scenario #2.

- The gradient-based approach introduced in Chapter 1 which is based on a PPNMM.
- The proposed RCA-SU algorithm which has been designed for the model in (4.1). It has been applied to this data set with $N_{\mathrm{MC}} = 3000$, $N_{\mathrm{bi}} = 2000$, $K = 4$ and $\beta = 1.2$.
- Finally, we consider the K-Hype method studied by Chen et al. (2013b) to compare our algorithm with state-of-the art kernel based unmixing methods. The kernel used in this study is the polynomial, second order symmetric kernel whose Gram matrix is defined by (4.12). This kernel provides better performance on this data set than the kernels studied by Chen et al. (2013b) (namely the Gaussian and the polynomial, second order asymmetric kernels). All hyperparameters of the K-Hype algorithm have been optimized using preliminary runs.

Table 4.4 compares the RNMSEs obtained with the SU algorithms for each class of the second scenario. These results show that the proposed algorithm provides abundance estimates similar to those obtained with the LMM-based algorithm (FCLS) for linearly mixed pixels. Moreover, the RCA-SU also provides accurate estimates for

the three mixing models considered, which illustrates the robustness of the RCA-based model regarding model mis-specification.

Table 4.4: Abundance RNMSEs ($\times 10^{-2}$): Scenario #2 .

Unmixing algo.	Class #0 (LMM)	Class #1 (GBM)	Class #2 (PPNMM)	Class #3 (RCA)
FCLS	**0.35**	9.20	19.74	30.73
GBM	**0.36**	3.05	15.24	29.53
PPNMM	0.65	**1.37**	**0.48**	23.77
K-HYPE	3.24	3.28	3.14	**3.42**
RCA-SU	**0.35**	**1.58**	**2.14**	**3.41**

The unmixing quality is also evaluated by the reconstruction error (RE) defined as

$$\mathrm{RE}_k = \sqrt{\frac{1}{N_k L} \sum_{n \in \mathcal{I}_k} \|\hat{\mathbf{y}}_n - \mathbf{y}_n\|^2} \tag{4.31}$$

where $\mathbf{y}_n$ is the nth observation vector and $\hat{\mathbf{y}}_n$ its estimate. Table 4.5 compares the REs obtained for the different classes. This table shows the accuracy of the proposed model for fitting the observations. The REs obtained with the RCA-SU are similar for the four pixel classes. Moreover, the performance in terms of RE of the proposed algorithm are similar to the performance of the K-Hype algorithm.

Table 4.5: REs ($\times 10^{-2}$): Scenario #2.

Unmixing algo.	Class #0 (LMM)	Class #1 (GBM)	Class #2 (PPNMM)	Class #3 (RCA)
FCLS	**0.99**	2.17	1.33	**3.10**
GBM	1.00	1.12	4.41	10.98
PPNMM	**0.99**	**1.01**	**0.99**	3.80
K-HYPE	**0.98**	**0.98**	**0.98**	**0.98**
RCA-SU	1.00	**0.98**	**0.98**	**0.98**

From a reconstruction point of view, the K-Hype and RCA-SU algorithms provides similar results. However, the proposed algorithm also provides nonlinearity detection maps. The PPNMM and RCA-SU algorithms perform similarly in term of abundance estimation and allow both nonlinearities to be detected in each pixel. However, the nonlinearities can be analyzed more deeply using the RCA-SU, as will be shown in the next part.

Nonlinearity detection

The performance of the proposed algorithm for nonlinearity detection is compared to the detector studied in Chapter 3, which is coupled with the PPNMM-based SU procedure mentioned above. The probability of false alarm of the PPNMM-based detection has been set to PFA = 0.05. Figs. 4.5 (c) and (d) show the detection maps obtained with the two detectors. Both detectors are able to locate the nonlinearly mixed regions. However, the RCA-SU provides more homogeneous regions, due to the consideration of spatial structure through the MRF.

Moreover, the proposed algorithm provides information about the different levels of nonlinearity in the image thanks to the estimation of the hyperparameters s_k^2 associated with the different classes. In this simulation, we obtain $[\hat{s}_1^2, \hat{s}_2^2, \hat{s}_3^2] = [0.2, 1.4, 10] \times 10^{-2}$, showing that nonlinearities of class $\mathcal{C}_1$ are less severe than those of class $\mathcal{C}_2$ that are themselves weaker than those of class $\mathcal{C}_3$. The next section studies the performance of the proposed algorithm for a real hyperspectral image.

4.8 Simulations for a real hyperspectral image

4.8.1 Data set

The real image considered in this section is the Villelongue image considered in Chapters 1 and 2. A sub-image (of size 41×29 pixels) is chosen here to evaluate the proposed unmixing procedure and is depicted in Fig. 4.6. The scene is composed mainly of roof, road and grass pixels, resulting in $R = 3$ endmembers. The spectral signatures of these components have been extracted from the data using the N-FINDR algorithm (Winter, 1999) and are depicted in Fig. 4.7.

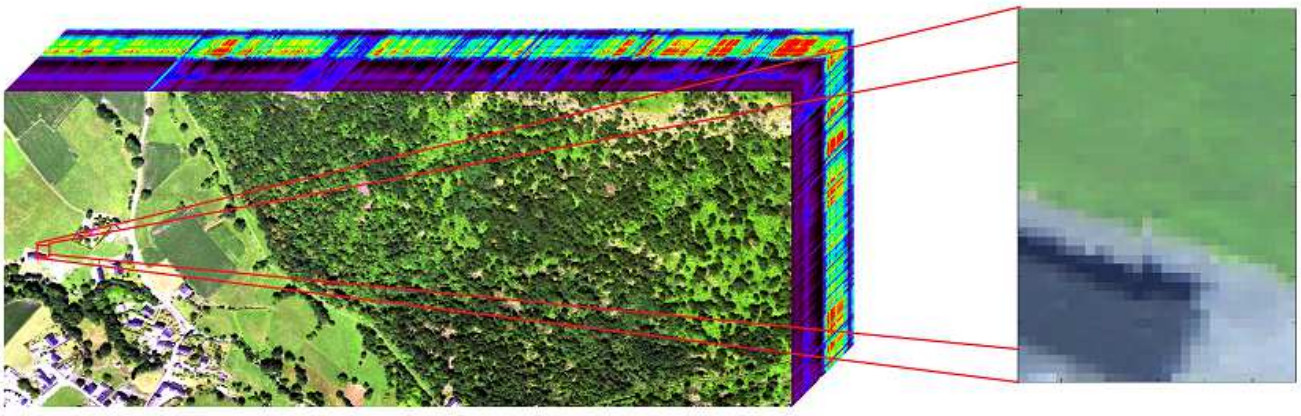


Figure 4.6: Real hyperspectral Madonna data acquired by the Hyspex hyperspectral scanner over Villelongue, France (left) and sub-image of interest (right).

4.8.2 Spectral unmixing

The proposed algorithm has been applied to this data set with $N_{\text{MC}} = 3000$ and $N_{\text{bi}} = 1000$. The number of classes has been set to $K = 4$ (one linear class and three nonlinear classes). The granularity parameter of the prior (4.14) has been fixed to $\beta = 0.7$. Fig. 4.8 shows examples of abundance maps estimated by the FCLS algorithm, the gradient-based method assuming the GBM, the PPNMM and the K-Hype (Chen et al., 2013b) algorithms and the proposed method. The abundance maps estimated by the RCA-SU algorithm are in good agreement with the state-of-the art algorithms. However, Table 4.6 shows that K-Hype and the proposed algorithm provide a lower reconstruction error. Fig. 4.9 compares the noise variances estimated by the RCA-SU for the real image with the noise variances estimated by the HySime algorithm (Bioucas-Dias and Nascimento, 2008). The HySime algorithm

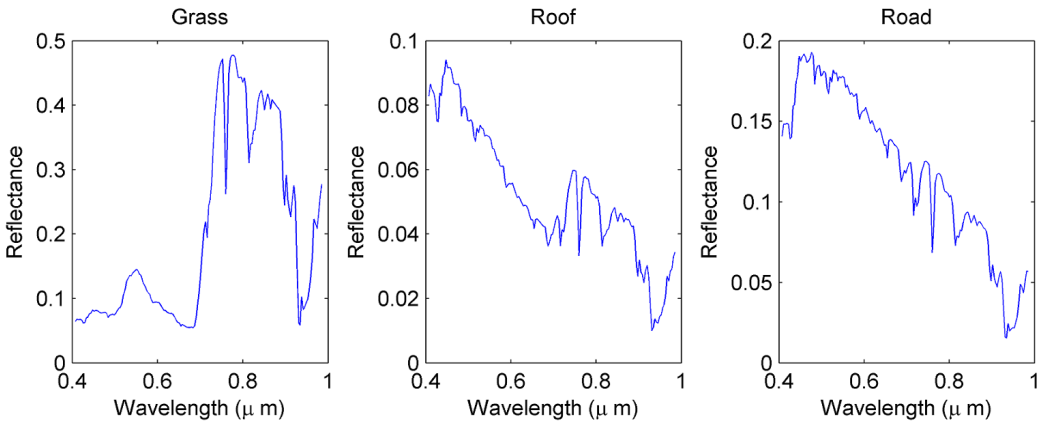


Figure 4.7: The $R = 3$ endmembers estimated by N-Findr for the real Madonna sub-image.

assumes additive noise and estimates the noise covariance matrix of the image using multiple regression. Fig. 4.9 shows that the two algorithms provide similar noise variance estimates. These results motivate the consideration of non i.i.d. noise for hyperspectral image analysis since the noise variances increase for the highest wavelengths. The simulations conducted on this real dataset show the accuracy of the proposed RCA-SU in terms of abundance estimation and reconstruction error, especially for applications where the noise variances vary depending on the wavelength. Moreover, it also provides information about the nonlinearities of the scene.

Table 4.6: Reconstruction errors: Real image.

Unmixing algo.	RE ($\times 10^{-2}$)
FCLS	0.65
GBM	0.65
PPNMM	**0.54**
K-HYPE	**0.48**
RCA-SU	**0.48**

4.8.3 Nonlinearity detection

Fig. 4.10 (b) shows the detection map (map of z_n for $n = 1, \ldots, N$) provided by the proposed RCA-SU detector for the real image considered. Due to the consideration of spatial structures, the proposed detector provides homogeneous regions. Similar structures can be identified in this detection map and the true color image of the scene (Fig. 4.10 (a)). The estimated class $\mathcal{C}_0$ (black pixels) associated with linearly mixed pixels is mainly located in the roof region. The class $\mathcal{C}_1$ (dark grey pixels) can be related to regions where the main component in the pixels are grass or road. Mixed pixels composed of grass and road are gathered in class $\mathcal{C}_2$ (light grey pixels). Finally, shadowed pixels located between the roof and the road are associated with the last class $\mathcal{C}_3$ (white pixels). Moreover, the RCA-SU can identify three levels of nonlinearity, corresponding to $[\hat{s}_1^2, \hat{s}_2^2, \hat{s}_3^2] = [0.03, 0.50, 29.5]$. The most influent nonlinearity class is class $\mathcal{C}_3$, where shadowing effects occurs. Mixed pixels of class $\mathcal{C}_2$ contain weaker nonlinearities. Finally, the remaining pixels of class $\mathcal{C}_1$ are associated with the weakest nonlinearities. The nonlinearities of this class

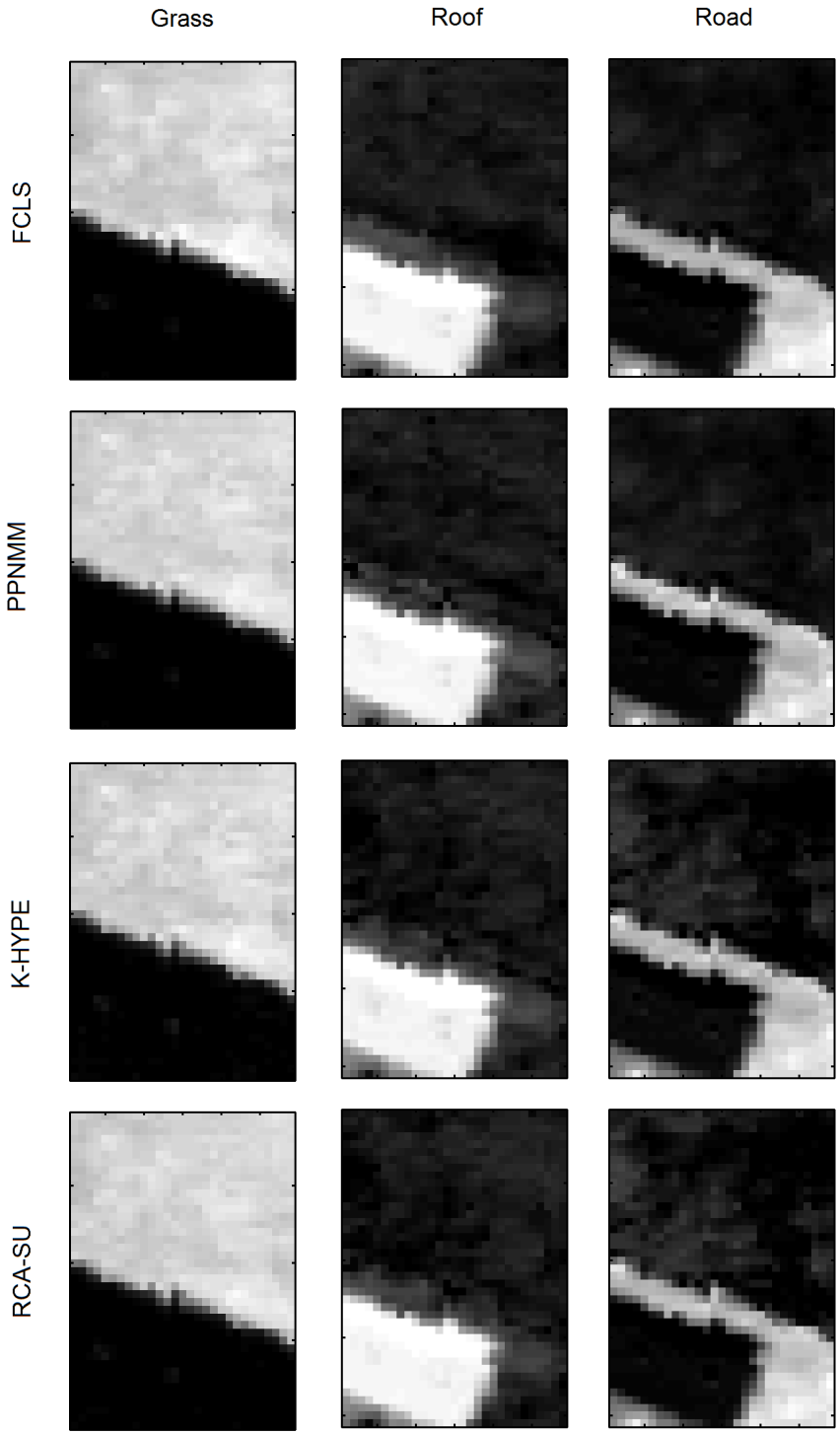


Figure 4.8: The $R = 3$ abundance maps estimated by the FCLS, PPNMM-based, K-Hype, and RCA-SU algorithms for the Madonna real image (white pixels correspond to large abundances, contrary to black pixels).

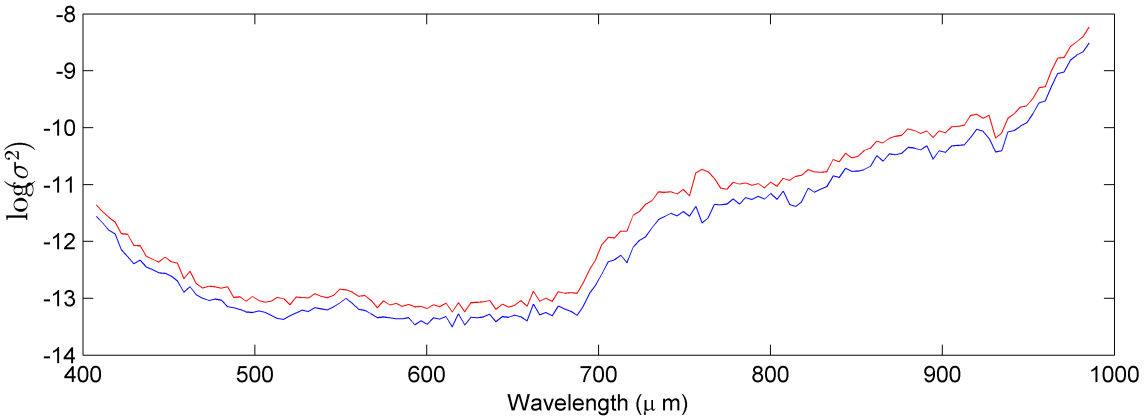


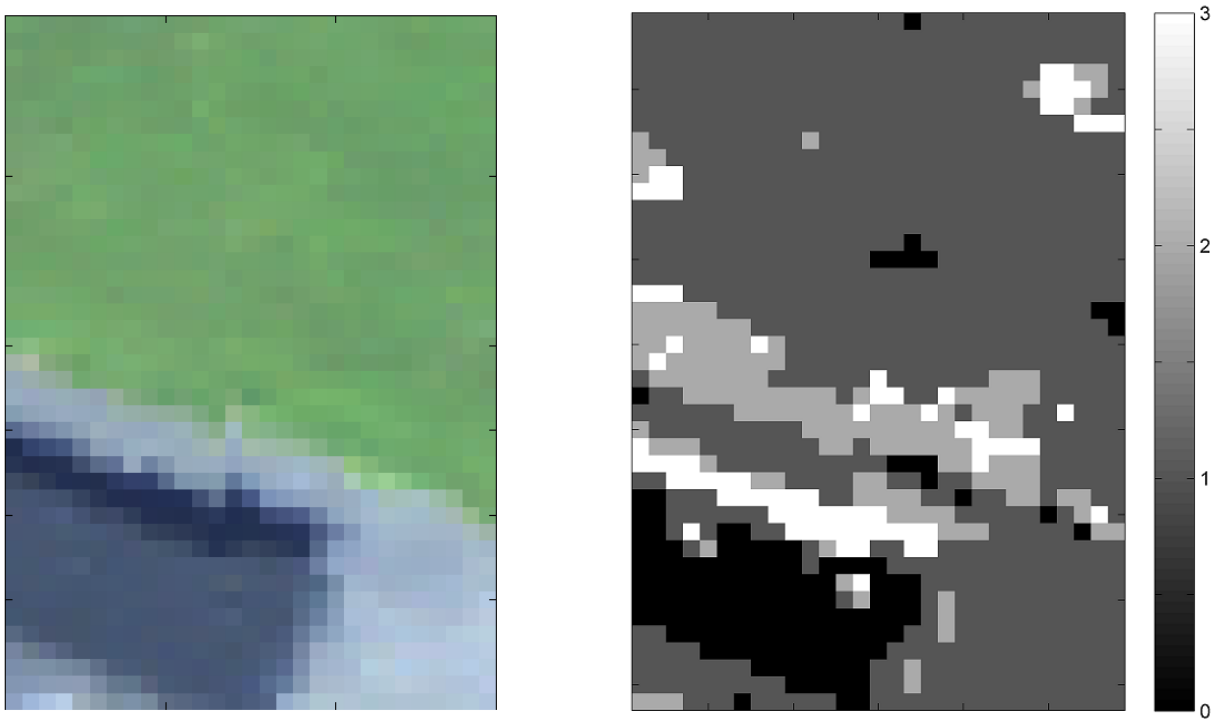
Figure 4.9: Noise variances estimated by the RCA-SU (red) and the Hysime algorithm (blue) for the real Madonna image.

can probably be explained by the endmember variability and/or the endmember estimation error. It is interesting to note that the RCA-SU identifies two rather linear classes associated with homogeneous regions mainly composed of a single parameter (classes $\mathcal{C}_0$ and $\mathcal{C}_1$). The two latter classes (classes $\mathcal{C}_2$ and $\mathcal{C}_3$) correspond to rather nonlinear regions where the pixels are mixed and shadowing effects occur.

4.9 Conclusion

We have proposed a new hierarchical Bayesian algorithm for joint linear/nonlinear spectral unmixing of hyperspectral images and nonlinearity detection. This algorithm assumed that each pixel of the image is a linear or nonlinear mixture of endmembers contaminated by additive Gaussian noise. The nonlinear mixtures are decomposed into a linear combination of the endmembers and an additive term representing the nonlinear effects. A Markov random field was introduced to promote spatial structures in the image. The image was decomposed into regions or classes where the nonlinearities share the same statistical properties, each class being associated with a level of nonlinearity. Nonlinearities within a same class were modeled using a Gaussian process parameterized by the endmembers and the nonlinearity level. Note finally that the physical constraints for the abundances were included in the Bayesian framework through appropriate prior distributions. Due to the complexity of the resulting joint posterior distribution, a Markov chain Monte Carlo method was investigated to compute Bayesian estimators of the unknown model parameters.

Simulations conducted on synthetic data illustrated the performance of the proposed algorithm for linear and nonlinear spectral unmixing. An important advantage of the proposed algorithm is its robustness regarding the actual underlying mixing model. Another interesting property resulting from the nonlinear mixing model considered is the possibility of detecting several kinds of linearly and nonlinearly mixed pixels. This detection can be used to identify the image regions affected by nonlinearities in order to characterize the nonlinear effects more deeply. Finally, simulations conducted with real data showed the accuracy of the proposed unmixing and nonlinearity



(a) (b)

Figure 4.10: (a) True color image of the scene of interest. (b) Nonlinearity detection map obtained with the RCA-SU detector for the Madonna image.

detection strategy for the analysis of real hyperspectral images.

As in Chapter 3, the endmembers contained in the hyperspectral image were assumed to be known in this work. Of course, the performance of the algorithm relies on this endmember knowledge. We think that estimating the pure component spectra present in the image, jointly with the abundance estimation and the nonlinearity detection is an important issue that should be considered in future work. Finally, the number of classes and the granularity of the scene were assumed to be known in this study. Estimating these parameters is clearly a challenging issue that should be investigated.

Main contributions. A new nonlinear mixing model for joint hyperspectral image unmixing and nonlinearity detection was proposed. The observed image was segmented into regions where nonlinear terms, if present, shared similar statistical properties. The resulting algorithm provided accurate abundance estimates when the actual mixtures are linear and nonlinear and it thus generalized the binary nonlinearity detectors proposed in the third chapter by considering different levels (classes) of nonlinearities.

4.10 Conclusion (in French)

Nous avons proposé un nouvel algorithme bayésien hiérarchique pour effectuer conjointement l'étape d'inversion et la détection de non-linéarités. Cet algorithme suppose que chaque pixel de l'image est un mélange linéaire ou non-linéaire des signatures spectrales des composants purs de l'image, contaminé par un bruit additif gaussien. Les mélanges non-linéaires sont décomposés en une combinaison linéaire des signatures spectrales des composants purs et d'un terme additif représentant les effets non-linéaires. Un champ de Potts-Markov a été introduit afin de promouvoir les structures spatiales dans l'image. L'image a été décomposée en régions (ou classes) où les non-linéarités ont les mêmes propriétés statistiques, chaque classe étant associée à un niveau de non-linéarité. Les non-linéarités dans une même classe ont été modélisés en utilisant un processus gaussien paramétré par les composants de l'image et un niveau de non-linéarité. Les contraintes physiques sur les abondances ont également été incluses dans le cadre bayésien à l'aide de lois *a priori* appropriées. En raison de la complexité de la loi *a posteriori* jointe résultante, une méthode MCMC a été utilisée pour calculer les estimateurs bayésiens des paramètres inconnus du modèle.
Les simulations effectuées sur des données synthétiques ont illustré les performances de l'algorithme proposé pour résoudre le problème de démélange spectral linéaire et non linéaire. Un avantage important de l'algorithme proposé est sa robustesse vis-à-vis du modèle de mélange réel sous-jacent. Une autre propriété intéressante résultant du modèle de mélange non-linéaire considéré est la possibilité de détecter plusieurs types de pixels résultant de mélanges non-linéaires. Cette détection peut être utilisée pour identifier les régions d'image affectées par des non-linéarités différentes dans le but de caractériser plus finement les effets non-linéaires. Enfin, les simulations effectuées avec des données réelles ont montré la pertinence des méthodes permettant d'effectuer conjointement l'étape d'inversion et la détection de non-linéarités pour l'analyse d'images réelles.
Comme dans le chapitre 3, les composantes spectrales pures contenues dans l'image hyperspectrale ont été supposées connues dans ce chapitre. Bien sûr, les performances de l'algorithme sont liées à la connaissance de ces spectres. Dans le futur, il serait important de proposer des méthodes permettant de résoudre conjointement le problème de démélange non-supervisé et la détection de non-linéarités. Enfin, le nombre de classes et la granularité de la scène ont également été supposés connus dans ce travail. L'estimation de ces paramètres est clairement une question difficile qui devrait être étudiée.
Contributions majeures. Un nouveau modèle de mélange non-linéaire pour effectuer conjointement l'étape d'inversion et la détection de non-linéarités a été proposée. L'image observée a été segmentée en régions où les termes non-linéaires, le cas échéant, partagent des propriétés statistiques similaires. L'algorithme résultant a fourni des abondances estimées satisfaisantes quand les mélanges réels sont linéaires et non linéaires et permet également de généraliser les détecteurs de non-linéarités binaires proposés dans le troisième chapitre de ce manuscrit en considérant différents niveaux (classes) de non-linéarités.

Conclusion and future work

The aim of this thesis was to study new nonlinear mixing models and to propose associated unmixing algorithms for hyperspectral image analysis. Spectral unmixing consists of extracting from a hyperspectral image, the spectra of the pure macroscopic components present in the image, referred to as endmembers, and of identifying their interactions referred to as mixtures. To reduce the problem complexity, most works of the literature rely on a linear mixing model, often considered as a first approximation of the actual mixture. However, this approximation can be inaccurate to describe some scenes, requiring more complex mixing models to be considered (to overcome the inherent limitations of the linear model). Designing and/or choosing an appropriate nonlinear mixing model for spectral unmixing is a challenging problem because of the diversity of nonlinear effects. Moreover, introducing nonlinear terms in the observation model complicates the derivation of efficient nonlinear unmixing procedures relying on nonlinear models. Considering nonlinear effects in hyperspectral images usually require more complex unmixing strategies than those assuming linear mixtures. Since the linear mixing model is often sufficient to approximate accurately most actual mixtures, it is also interesting to detect pixels or regions where the linear model is accurate.

4.11 Conclusion

The first chapter studied a post-nonlinear mixing model (PNMM) for nonlinear SU. The proposed polynomial PNMM (PPNMM) investigated in this chapter allowed nonlinearities to differ for each pixel, leading to a flexible mixture characterization. The nonlinearity of each pixel was characterized by a single parameter which is zero when the pixel is linearly mixed. First, supervised SU methods based on the PPNMM were considered. The first proposed algorithm was a hierarchical Bayesian algorithm coupled with MCMC methods. Two alternative optimization methods were also introduced to reduce the computational complexity of the sampling algorithm. The flexibility of the PPNMM and the performance of the three methods were evaluated using simulations conducted on synthetic and real data. The second part of this chapter presented a new Bayesian model for unsupervised SU based on the PPNMM. Appropriate priors were also assigned to the unknown endmembers to be sampled. The joint estimation of the abundances and endmembers required a large number of parameters to be sampled. To improve the mixing properties of the sampler, constrained Hamiltonian Monte Carlo methods were investigated.
The second chapter of this manuscript considered a kernel-based approach for nonlinear SU based on a nonlinear dimensionality reduction using a GPLVM. GPLVMs, which have received growing interest in the machine learning community, have the ability to approximate various nonlinear mappings from a low-dimensional space (latent space) to a higher dimensional observation space through the use of kernel functions, which makes them particularly well

adapted for hyperspectral image analysis and thus unmixing. In this chapter, we proposed to use a particular form of kernel based on existing bilinear models, which allowed the proposed unmixing strategy to be accurate when the underlying mixing model is bilinear. The proposed unmixing algorithm breaks the usual paradigm of spectral unmixing by first estimating the abundances and then predicting the endmembers using GPs. Simulations conducted on synthetic and real images illustrated the flexibility of the proposed model for linear and nonlinear spectral unmixing and provided promising results for abundance and endmember estimations even in the absence of pure pixels in the image.

The third chapter of this manuscript was dedicated to the detection of nonlinearities in hyperspectral images. This chapter focused on supervised detection procedures, i.e., the endmembers were assumed to be *a priori* known. The first detector was based on the PPNMM studied in the first chapter. More precisely, the associated test was based on the statistical properties of the PPNMM parameters to decide whether a given pixel is linearly or nonlinearly mixed. Conversely, the second test proposed did not assume any particular NLMM, i.e., only relied on the LMM. The two detectors were compared using simulations conducted on synthetic and real data.

In the last chapter of this manuscript, a new nonlinear mixing model for joint hyperspectral image unmixing and nonlinearity detection was proposed. The observed image was segmented into regions where nonlinear terms, if present, shared similar statistical properties. The resulting algorithm provided accurate abundance estimates for linear and nonlinear mixtures. Thus, it generalized the binary nonlinearity detectors proposed in the third chapter by considering different classes (levels) of nonlinearities.

In this manuscript, it was shown that nonlinear unmixing algorithms can improve the characterization of hyperspectral images compared to methods based on a linear model. These methods allowed the reconstruction errors to be reduced. Moreover, these methods provided better spectral signature and abundance estimates when the observed pixels result from nonlinear mixtures. The simulation results conducted on synthetic and real images illustrated the advantage of using nonlinearity detectors for hyperspectral image analysis. In particular, the proposed detectors can identify components which are present in few pixels (and hardly distinguishable) and can locate areas where significant nonlinear effects occur (shadow, relief, ...).

4.12 Future work

In this manuscript, we proposed nonlinear unmixing and nonlinearity detection algorithms using a polynomial post-nonlinear model. The flexibility of the Bayesian framework allowed the model complexity to be handled and prior information about the model parameters to be easily included within the estimation procedure. It would be interesting to extend this work to physics-based nonlinear models of the literature such as polynomial (bilinear) (Halimi et al., 2011a; Meganem et al., 2013) and intimate mixture models (Hapke, 1981).

Consider the following bilinear model

$$\mathbf{y}_n = \sum_{r=1}^{R} a_{r,n}\mathbf{m}_r + \sum_{i=1}^{R-1}\sum_{j=i+1}^{R} \beta_{i,j,n}\mathbf{m}_i \odot \mathbf{m}_j + \mathbf{e}_n \tag{4.32}$$

defined in (3). As mentioned in Chapter 4, the interaction spectra $\mathbf{m}_i \odot \mathbf{m}_j$ are highly correlated which complicates the estimation of the nonlinearity parameters $\beta_{i,j,n}$, especially for large numbers of endmembers R. In Chapter 4, we considered such parameters as nuisance parameters which were marginalized. However, if one is particularly

interested in these parameters (to quantify the influence of each interaction), additional information should be included in the estimation procedure. As mentioned by Gader et al. (2012), most of the $\beta_{i,j,n}$ parameters are expected to be small (or even null). Consequently, it is interesting to enforce small values and/or sparsity for these unknown parameters. In the Bayesian framework, sparsity promoting prior distributions (such as Laplace, Bernoulli-Gaussian or Bernoulli-Laplace distributions) could be used to infer the nonlinearity parameters.

Even if the consideration of nonlinear terms increases the unmixing complexity, the recent advances in efficient simulation methods (including Hamiltonian Monte Carlo methods) for high dimensional problems are promising and their application to hyperspectral analysis is an interesting prospect.

The second chapter of this manuscript focused on a kernel-based nonlinear dimensionality reduction technique initially studied in the machine learning community. A particular nonlinear mapping adapted for modeling bilinear nonlinearities was used and provided encouraging results. Moreover, the fourth chapter of this manuscript focused on a kernel-based model for joint unmixing and nonlinearity detection. Kernel-based methods for hyperspectral unmixing have recently received growing interest (Broadwater and Banerjee, 2009; Chen et al., 2013b) for their flexibility. However, further efforts should be achieved for proposing more general but accurate kernel functions. In the fourth chapter of this manuscript, the nonlinearity of each pixel was characterized by a kernel function which only depends on the endmember matrix. However, it makes sense to assume that the nonlinearity may also depend on the pixel composition, i.e., on its abundances. Kernels depending on the endmembers and the abundances, such as in the kernel-based method recently proposed by Chen et al. (2013a), should be investigated.

In the two last chapters of the manuscript, the endmembers were assumed to be known to perform the nonlinearity detection and the joint unmixing and nonlinearity detection. However, if the endmembers are unknown, they have to be extracted from the data, which can be difficult when nonlinearly mixed pixels are present in the scene. It would be interesting to apply nonlinearity detectors before the endmember extraction (and then using linear EEAs to identify the pure spectral signatures) or to propose joint nonlinearity detection and endmember estimation methods. Thus, it would be interesting to extend the RCA-based Bayesian model proposed in Chapter 4 to the case where the endmembers are included in the estimation procedure.

In this manuscript, we considered linear and nonlinear mixing models whose parameters were estimated using Bayesian inference. Thus, the dimension of the unknown parameter vectors was known and thus their admissible sets were assumed to be unique. In particular, the number of endmembers R was fixed from *a priori* knowledge. In the RCA-based model, the number of classes was also fixed. If the dimension of the unknown parameter vector is unknown, a simple solution consists of running the proposed algorithms with different numbers of endmembers and/or classes. However, the computational complexity of this approach can be prohibitive when the number of parameters to be tuned is large. A more elegant approach consists of including the dimension of the unknown parameter vector within the parameter estimation using the Bayesian framework. Sampling from the resulting posterior distribution can be achieved using reversible jump MCMC (RJ-MCMC) methods Green (1995). RJ-MCMC methods can be used for model order selection but also to sample variables that can live in disjoint sets, which is particularly interesting for hyperspectral image analysis. For instance, prior distributions defined on a set of (disjoint) domains could be assigned to unknown endmembers and/or abundances. As a toy example, consider the problem of semi-supervised unmixing of a hyperspectral image that consists of urban and forested areas. The problem is semi-supervised in the sense that the endmembers present in the image belong to a spectral library which contains spectra of both urban and forested areas. Instead of searching the active endmembers in the whole library,

it would be interesting to identify the urban and forested areas in the scene and to estimate the endmembers out of subsets of the library. For this problem, prior distributions for the unknown endmembers could be defined on different subsets of the library and RJ-MCMCs could be used to sample for the resulting posterior.

This example is highly related to the sparse linear unmixing problem that have received intensive interest over the last few years in the hyperspectral community ((Bioucas-Dias et al., 2012) and references therein). This sparse linear regression problem has strong relations with compressed sensing and basis pursuit. It consists of solving the inversion step given a large spectral library by enforcing the abundances to be sparse. Optimization algorithms are mainly used to solve this problem and the ℓ_1 norm penalization is often preferred (ℓ_1 relaxation) rather than the ℓ_0 norm, which is more challenging to handle. Conversely, the Bayesian framework is flexible enough to handle penalizations based on the ℓ_0 norm, using RJ-MCMCs for instance. Moreover, this framework allows also for structured sparsity to be considered, such as in (Dobigeon and Févotte, 2013; Iordache et al., 2013). Proposing Bayesian sparse unmixing procedures is a very interesting prospect. However, efficient sampling procedures must be proposed to handle the complexity induced by the consideration of prior sparsity information.

Finally, consideration of endmember variability for spectral unmixing is also interesting. In this manuscript, the spectral signatures of the materials of the scene were assumed to be unique. However, the spectral variability of these signatures cannot be neglected in some images. Solving the unmixing problem in such cases is challenging and should deserve deeper attention in future work.

Conclusion et perspectives (in French)

L'objectif de cette thèse était d'étudier de nouveaux modèles de mélange non-linéaires et de proposer des algorithmes de démélange associés pour l'analyse d'images hyperspectrales. Le problème de démélange spectral consiste à extraire d'une image hyperspectrale, les spectres des composants macroscopiques purs présents dans l'image, et à idenfifier leurs interactions ou mélanges. Pour réduire la complexité du problème, la plupart des travaux de la littérature s'appuient sur un modèle de mélange linéaire, souvent considéré comme une première approximation des mélanges réels. Cependant, cette approximation peut ne pas être appropriée pour décrire certaines scènes qui nécessitent d'utiliser des modèles de mélange plus complexes. Définir et/ou choisir un modèle de mélange non-linéaire approprié pour le démélange spectral est un problème difficile en raison de la diversité des effets non-linéaires. De plus, l'introduction de termes non-linéaires dans le modèle d'observation complique la mise en place de procédures de démélange efficaces reposant sur des modèles non-linéaires. Considérer des effets non-linéaires dans les images hyperspectrales exige habituellement des algorithmes de démélange plus complexes que ceux basés sur un modèle linéaire. Étant donné que le modèle de mélange linéaire est souvent suffisant pour modéliser avec précision la plupart des mélanges réels, il est aussi intéressant de détecter les pixels ou les régions où le modèle linéaire peut suffire.

4.13 Conclusion

Dans le premier chapitre, nous avons étudié un modèle de mélange post-non-linéaire (PNMM) pour résoudre le problème de démélange non-linéaire. Le modèle proposé (PPNMM) étudié dans ce chapitre permet de considérer des non-linéarités différentes pour les différents pixels de l'image. La non-linéarité de chaque pixel est caractérisée par un paramètre unique qui est égal à zéro lorsque le pixel résulte d'un mélange linéaire. Tout d'abord, des méthodes de démélange supervisé basées sur le PPNMM ont été présentées. Le premier algorithme proposé est un algorithme bayésien hiérarchique couplée avec des méthodes MCMC. Deux méthodes d'optimisation alternatives ont également été introduites afin de réduire le coût calculatoire de l'échantillonneur de Gibbs. La souplesse du PPNMM et les performances des trois méthodes ont été évaluées à l'aide de simulations effectuées sur données synthétiques et réelles. La deuxième partie de ce chapitre a présenté un nouveau modèle bayésien pour résoudre le problème de démélange non-supervisé. Des lois a priori appropriées ont été également attribuées aux signatures spectrales inconnues et à échantillonner. L'estimation conjointe des abondances et de ces signatures spectrales implique un grand nombre de paramètres à échantillonner. Pour améliorer les propriétés de mélange de l'échantillonneur, des méthodes de Monte Carlo à dynamiques hamiltoniennes contraintes ont été étudiées.
Le second chapitre de ce manuscrit s'est focalisé sur une méthode à noyaux basée sur une réduction de dimension

non-linéaire utilisant un modèle à variables latentes et des processus gaussiens (GPLVM). Les GPLVMs, qui ont suscité un intérêt croissant dans la communauté de l'apprentissage automatique ont la capacité d'approcher à grand nombre d'applications non-linéaires d'un espace de faible dimension (espace latent) à un espace d'observation de dimension supérieure grâce à l'utilisation de noyaux, ce qui les rend particulièrement bien adaptés pour l'analyse d'images hyperspectrales et donc pour le démélange spectral. Dans ce chapitre, nous avons proposé d'utiliser une forme particulière de noyau basé sur des modèles bilinéaires existants, ce qui a permis d'obtenir un algorithme de démélange bien adapté lorsque le modèle de mélange sous-jacent est bilinéaire. L'algorithme de démélange proposée rompt le paradigme habituel du démélange spectral en estimant tout d'abord les abondances puis en prédisant les spectres de composants purs à l'aide de processus gaussiens. Les simulations effectuées sur des images synthétiques et réelles ont illustré la flexibilité du modèle proposé pour le démélange linéaire et non-linéaire et ont donné des résultats prometteurs, même en l'absence de pixels purs dans l'image observée.
Le troisième chapitre de ce manuscrit a été consacrée à la détection de non-linéarités dans les images hyperspectrales. Ce chapitre s'est concentré sur des procédures de détection supervisées dans le sens où les spectres des composants présents dans l'image étaient supposés connus. Le premier détecteur était fondé sur le modèle PPNMM étudié dans le premier chapitre. Plus précisément, le test proposé était basé sur les propriétés statistiques de l'estimateur du maximum de vraisemblance des paramètres du modèle PPNMM afin de décider si un pixel résulte d'un mélange linéaire où non-linéaire. A contrario, le deuxième détecteur proposé ne supposait pas de formes de non-linéarités particulière, c'est-à-dire, ne reposait que sur le modèle de mélange linéaire. Les deux détecteurs ont été comparés à l'aide de simulations effectuées sur données synthétiques et réelles.
Dans le dernier chapitre de ce manuscrit, un nouveau modèle de mélange non-linéaire pour effectuer conjointement le démélange supervisé et la détection de non-linéarités a été proposé. L'image observée a été segmentée en régions où les termes non-linéaires (lorsqu'ils sont présents) partagent des propriétés statistiques similaires. L'algorithme résultant a montré de bonnes performances en terme d'estimation des abondances lorsque les pixels de l'image résultent de mélange linéaires et non-linéaires. De plus, il généralise les détecteurs de non-linéarités binaires proposés dans le troisième chapitre en considérant différentes classes (niveaux) de non-linéarités.
Dans ce manuscrit, il a été montré que les algorithmes de démélange non-linéaires peuvent permettre d'améliorer la caractérisation des images hyperspectrales par rapport aux méthodes basées sur un modèle linéaire. Ces méthodes non-linéaires ont montré de bonnes performances en termes d'estimation des abondances lorsque les pixels de l'image résultent de mélange non-linéaires. Les résultats de simulations effectuées sur des images synthétiques et réelles ont illustré l'avantage de l'utilisation de détecteurs de non-linéarités pour l'analyse d'images hyperspectrales. En particulier, les détecteurs proposés permettent d'identifier des composants très peu représentés (et difficile à détecter) et peuvent localiser des régions où se produisent des effets non-linéaires importants (ombre, relief, ...).

4.14 Perspectives

Dans ce manuscrit, nous avons proposé des algorithmes de détection de non-linéarités et de démélange non-linéaires en utilisant un modèle post-non-linéaire polynômial. La flexibilité du cadre bayésien a permis de gérer la complexité du modèle à traiter et a également permis d'inclure facilement l'information disponible sur les paramètres du modèle dans la procédure d'estimation. Il serait intéressant d'étendre ce travail à d'autres modèles non-linéaires de la littérature, en particulier ceux motivés par les phénomènes physiques en jeu dans les processus de mélange (tels que

les modèles polynômiaux (Halimi et al., 2011a; Meganem et al., 2013) et les modèles de mélanges intimes (Hapke, 1981)).

Considérons le modèle bilinéaire suivant

$$\mathbf{y}_n = \sum_{r=1}^{R} a_{r,n}\mathbf{m}_r + \sum_{i=1}^{R-1}\sum_{j=i+1}^{R} \beta_{i,j,n}\mathbf{m}_i \odot \mathbf{m}_j + \mathbf{e}_n \tag{4.33}$$

défini par l'Eq. (3). Comme noté dans le chapitre 4, les spectres d'interactions $\mathbf{m}_i \odot \mathbf{m}_j$ sont fortement corrélés, ce qui complique l'estimation des paramètres de non-linéarité $\beta_{i,j,n}$, en particulier pour un grand nombre de composants R. Dans le chapitre 4, nous avons considéré ces paramètres comme des paramètres de nuisance qui ont été marginalisés. Toutefois, si l'on est particulièrement intéressé par les valeurs de ces paramètres (si l'on veut quantifier l'influence de chaque interaction), de l'information supplémentaire doit être incluse dans la procédure d'estimation. Comme noté par Gader et al. (2012), on peut s'attendre à ce que les paramètres $\beta_{i,j,n}$ soient faibles (ou même nuls). Par conséquent , il serait intéressant de promouvoir les faibles valeurs (voire la parcimonie) pour ces paramètres. Dans le cadre bayésien, des lois *a priori* favorisant la parcimonie (par exemple, des lois Laplace, Bernoulli-gaussienne ou Bernoulli-Laplace) pourraient être utilisées pour estimer les paramètres de non-linéarité.

Même si la considération de termes non-linéaires augmente la complexité du problème de démélange, les récents progrès dans le domaine des méthodes de simulation (y compris les méthodes de Monte Carlo à dynamiques hamiltoniennes) pour des problèmes de grande dimension sont prometteurs. Leur application à l'analyse d'images hyperspectrales est une perspective intéressante.

Le deuxième chapitre de ce manuscrit a porté sur une technique de réduction de dimension à base de noyau étudiée à l'origine par la communauté de l'apprentissage automatique. Un noyau particulier et adapté aux non-linéarités bilinéaires a été utilisé et a fourni des résultats encourageants. De plus, le quatrième chapitre de ce manuscrit s'est focalisé sur un modèle à noyau pour effectuer conjointement le démélange et la détection de non-linéarités. Les méthodes à noyaux pour le démélange spectral ont récemment reçu un intérêt croissant, (Broadwater and Banerjee, 2009; Chen et al., 2013b) motivé par leur grande flexibilité. Toutefois, des efforts supplémentaires doivent être réalisés pour proposer des noyaux plus généraux mais toujours adaptés à l'imagerie hyperspectrale. Dans le quatrième chapitre de ce manuscrit, la non-linéarité de chaque pixel était caractérisée par une fonction de noyau qui ne dépendait que de la matrice des signatures spectrales. Toutefois, il est raisonnable de penser que la non-linéarité peut également dépendre de la composition de pixel, c'est-à-dire, de ses abondances. La modélisation des non-linéarités en utilisant les signatures spectrales pures et les abondances, comme dans la méthode récemment proposé par Chen et al. (2013a), est une perspective intéressante.

Dans les deux derniers chapitres du manuscrit, les signatures spectrales des composants de l'image étaient supposées connues pour pourvoir effectuer la détection de non-linéarités et le démélange non-linéaire. Toutefois, si ces spectres sont inconnus, ils doivent être extraits des données, ce qui peut être difficile lorsque des pixels résultant de mélanges non-linéaires sont présents dans la scène. Il serait intéressant d'appliquer des détecteurs de non-linéarités avant l'extraction de ces signatures ou de proposer des méthodes réalisant conjointement l'estimation de ces spectres et la détection de non-linéarités. Par exemple, il serait intéressant d'étendre le modèle bayésien basé sur RCA proposé au chapitre 4 pour le cas où les signatures spectrales pures sont incluses dans la procédure d'estimation.

Dans ce manuscrit, nous avons examiné des modèles de mélange linéaires et non-linéaires dont les paramètres étaient estimés par inférence bayésienne. De plus, la dimension des vecteurs de paramètres inconnus était connue et leurs ensembles de définition étaient uniques. En particulier, le nombre de composants R était fixé *a priori*. Dans

le modèle inspiré de RCA, le nombre de classes était également fixé. Si la dimension du vecteur de paramètres inconnu est inconnue, une solution simple consiste à exécuter les algorithmes proposés avec différents nombres de composants et/ou de classes. Cependant, le coût calculatoire de cette approche peut être prohibitif lorsque le nombre de paramètres à tester est important. Une approche plus élégante consiste à inclure la dimension du vecteur de paramètres inconnu dans la procédure d'estimation en utilisant le cadre bayésien. L'échantillonnage de la loi *a posteriori* qui en résulte peut être réalisé en utilisant des MCMC à sauts réversibles (RJ-MCMC) Green (1995). Les méthodes de type RJ-MCMC peuvent être utilisées pour sélectionner l'ordre d'un modèle, mais aussi pour échantillonner des paramètres qui peuvent vivre dans des ensembles disjoints, ce qui est particulièrement intéressant pour l'analyse d'images hyperspectrales. Par exemple, des lois *a priori* définies sur un ensemble de domaines (disjoints) peuvent être utilisées pour les signatures spectrales et/ou les abondances à estimer. Un exemple d'application pourrait être le problème de démélange semi-supervisé d'une image hyperspectrale qui se compose de zones urbaines et boisées. Le problème est semi-supervisé dans le sens où les composants présents dans l'image appartiennent à une bibliothèque spectrale qui contient les spectres des deux zones (urbaines et boisées). Au lieu de chercher les signatures spectrales actives dans l'ensemble de la bibliothèque, il serait intéressant d'identifier les zones urbaines et boisées de la scène et de rechercher les signatures spectrales dans des sous-ensembles de la bibliothèque. Pour ce problème, des lois *a priori* pour les signatures spectrales inconnues pourraient être définies sur différents sous-ensembles de la bibliothèque et des méthodes de type RJ-MCMC pourraient être utilisées pour échantillonner la loi *a posteriori* associée au modèle bayésien.

Cet exemple est fortement liée au problème de démélange linéaire parcimonieux qui a suscité un fort intérêt au cours des dernières années dans la communauté hyperspectrale (Bioucas-Dias et al., 2012). Ce problème consiste à résoudre l'étape d'inversion (estimation des abondances) à partir d'une large bibliothèque spectrale en forçant les abondances à être parcimonieuses. Des algorithmes d'optimisation sont souvent utilisés pour résoudre ce problème et la pénalisation par la norme ℓ_1 est souvent préférée (relaxation par norme ℓ_1) à la norme ℓ_0 qui est plus difficile à manipuler. Inversement, le cadre bayésien est suffisamment souple pour gérer des pénalisations utilisant la norme ℓ_0, en utilisant des méthodes de type RJ- MCMC par exemple. De plus, ce cadre permet également de considérer des parcimonies structurées, comme dans (Dobigeon and Févotte, 2013; Iordache et al., 2013). Proposer les algorithmes bayésien de démélange parcimonieux est une perspective très intéressante. Toutefois, des échantillonneurs efficaces doivent être proposés pour gérer la complexité induite par la prise en compte de parcimonies.

Enfin, l'étude de la variabilité de signatures spectrales pour le démélange est également intéressante. Dans ce manuscrit, les signatures spectrales des matériaux de la scène étaient supposées être uniques. Cependant, la variabilité spectrale de ces signatures ne peut pas être négligée dans certaines images. Résoudre le problème de démélange dans de telles situations est difficile et ce cas devrait être approfondi dans le futur.

Appendices

Appendix A

Identifiability of the supervised PPNMM-based SU problem

A.1 Non-injectivity of $\mathbf{s} \mapsto \mathbf{g}(\mathbf{s})$

The application

$$\begin{aligned} \mathbf{g}: \quad [0,1]^L &\to \mathbb{R}^L \\ \mathbf{s} &\mapsto \left[s_1 + bs_1^2, \ldots, s_L + bs_L^2\right]^T \end{aligned}$$

with $\mathbf{s} = [s_1, \ldots, s_L]^T$ is not injective. Indeed, if $\mathbf{g}(\mathbf{s}_1) = \mathbf{g}(\mathbf{s}_2)$ then

$$\forall l = 1, \ldots, L \quad s_{2,i} \in \left\{ s_{1,i}, -\frac{1}{b} - s_{1,i} \right\}$$

which leads to 2^L solutions for the problem $\mathbf{g}(\mathbf{s}_1) = \mathbf{g}(\mathbf{s}_2)$

A.2 Injectivity of $(\boldsymbol{a}, b) \mapsto \mathbf{g}(\boldsymbol{a}, b) = \mathbf{M}\boldsymbol{a} + b(\mathbf{M}\boldsymbol{a}) \odot (\mathbf{M}\boldsymbol{a})$

Let $\boldsymbol{a}$ and $\boldsymbol{a}^*$ be two abundance vectors satisfying the positivity and sum-to-one constraints, $\mathbf{M}$ the matrix containing the endmembers and $(b, b^*) \in \mathbb{R}^2$. Consider the nonlinear functional $\mathbf{g}(\boldsymbol{a}, b)$ defined as follows

$$\begin{aligned} \mathbf{g}(\boldsymbol{a}, b) &= \mathbf{M}\boldsymbol{a} + b(\mathbf{M}\boldsymbol{a}) \odot (\mathbf{M}\boldsymbol{a}) \\ &= \sum_{r=1}^{R} a_r \mathbf{m}_r + b \sum_{r=1}^{R} a_r^2 \mathbf{m}_r \odot \mathbf{m}_r + 2b \sum_{r=1}^{R=1} \sum_{j=r+1}^{R} a_r a_j \mathbf{m}_r \odot \mathbf{m}_j \end{aligned}$$

where $\odot$ denotes the term by term product operation. If $\mathbf{g}(\boldsymbol{a}, b) = \mathbf{g}(\boldsymbol{a}^*, b^*)$, then

$$\mathbf{M}(\boldsymbol{a} - \boldsymbol{a}_2^*) + b(\mathbf{M}\boldsymbol{a}) \odot (\mathbf{M}\boldsymbol{a}) - b^*(\mathbf{M}\boldsymbol{a}^*) \odot (\mathbf{M}\boldsymbol{a}^*) = 0$$

and

$$\sum_{r=1}^{R} (a_r - a_r^*)\mathbf{m}_r + \sum_{r=1}^{R} (ba_r^2 - b^* a_r^{*2})\mathbf{m}_r \odot \mathbf{m}_r + 2 \sum_{r=1}^{R=1} \sum_{j=r+1}^{R} (ba_r a_j - b^* a_r^* a_j^*)\mathbf{m}_r \odot \mathbf{m}_j = 0.$$

If the columns of the $L \times \frac{R(R+3)}{2}$ matrix

$$\check{\mathbf{M}} = \{\mathbf{m}_1, \ldots, \mathbf{m}_R, \mathbf{m}_1 \odot \mathbf{m}_1, \ldots, \mathbf{m}_R \odot \mathbf{m}_R, \mathbf{m}_1 \odot \mathbf{m}_2, \ldots, \mathbf{m}_{R-1} \odot \mathbf{m}_R\}$$

are linearly independent, then

$$\boldsymbol{a} = \boldsymbol{a}^*, \quad \text{and} \quad b = b^*. \tag{A.1}$$

Consequently, the identifiability of the unmixing problem assuming the proposed PPNMM is ensured when $\text{rank}(\check{\mathbf{M}}) = \frac{R(R+3)}{2}$, which is usually satisfied when using real pure spectral components. Note that the identifiability of the unmixing problem associated with the LMM requires a similar condition, i.e., $\text{rank}(\mathbf{M}) = R$, where $\mathbf{M} = [\mathbf{m}_1, \ldots, \mathbf{m}_R]^T$.

Appendix B

Partial derivatives for the LS PPNMM-based algorithms

B.1 Partial derivatives of $\check{b}(\cdot)$ and $\mathbf{h}(\cdot)$

The partial derivative of $\mathbf{h}(\cdot)$ with respect to $a_r, (r = 1, \ldots, R)$ is given by

$$\frac{\partial \mathbf{h}(\boldsymbol{a})}{\partial a_r} = 2\,(\mathbf{M}\boldsymbol{a}) \odot \mathbf{m}_r.$$

Using the following partial derivatives

$$\begin{aligned}
\frac{\partial(\mathbf{y} - \mathbf{M}\boldsymbol{a})}{\partial a_r} &= -\mathbf{m}_r \\
\frac{\partial(\mathbf{y} - \mathbf{M}\boldsymbol{a})^T\mathbf{h}(\boldsymbol{a})}{\partial a_r} &= -\mathbf{m}_r^T\mathbf{h}(\boldsymbol{a}) + 2(\mathbf{y} - \mathbf{M}\boldsymbol{a})^T\left((\mathbf{M}\boldsymbol{a}) \odot \mathbf{m}_r\right) \\
\frac{\partial \mathbf{h}(\boldsymbol{a})^T\mathbf{h}(\boldsymbol{a})}{\partial a_r} &= 2\mathbf{h}(\boldsymbol{a})^T\frac{\partial \mathbf{h}(\boldsymbol{a})}{\partial a_r}
\end{aligned}$$

and the usual differentiation rules, we obtain

$$\frac{\partial \check{b}(\boldsymbol{a})}{\partial a_r} = \frac{1}{\|\mathbf{h}(\boldsymbol{a})\|^4}\left[\frac{\partial(\mathbf{y} - \mathbf{M}\boldsymbol{a})^T\mathbf{h}(\boldsymbol{a})}{\partial a_r}\mathbf{h}(\boldsymbol{a})^T\mathbf{h}(\boldsymbol{a}) - (\mathbf{y} - \mathbf{M}\boldsymbol{a})^T\mathbf{h}(\boldsymbol{a})\frac{\partial \mathbf{h}(\boldsymbol{a})^T\mathbf{h}(\boldsymbol{a})}{\partial a_r}\right].$$

B.2 Partial derivatives of $\check{\mathbf{y}}^*(\cdot)$

Thanks to the sum-to-one constraint of the abundance vector, the cost function (1.23) can be expressed as a function of $\mathbf{c} = [a_1, \ldots, a_{R-1}]^T$ by setting $a_R = 1 - \sum_{r=1}^{R-1} a_r$. Straightforward computations lead to

$$\mathbf{M}\boldsymbol{a} = \overline{\mathbf{M}}\mathbf{c} + \mathbf{m}_R$$

where $\overline{\mathbf{M}} \triangleq [\mathbf{m}_1 - \mathbf{m}_R, \ldots, \mathbf{m}_{R-1} - \mathbf{m}_R] \triangleq [\overline{\mathbf{m}}_1, \ldots, \overline{\mathbf{m}}_{R-1}]$ is a matrix of size $L \times (R-1)$. The spectrum $\check{\mathbf{y}}(\mathbf{c})$ can then be expressed as

$$\check{\mathbf{y}}(\mathbf{c}) = \overline{\mathbf{M}}\mathbf{c} + \mathbf{m}_R + \check{b}^*(\mathbf{c})\overline{\mathbf{h}}(\mathbf{c})$$

where

$$\check{b}^*(\mathbf{c}) = \check{b}\left(a_1, \ldots, a_{R-1}, 1 - \sum_{r=1}^{R-1} a_r\right)$$

and

$$\overline{\mathbf{h}}(\mathbf{c}) = (\overline{\mathbf{M}}\mathbf{c} + \mathbf{m}_R) \odot (\overline{\mathbf{M}}\mathbf{c} + \mathbf{m}_R).$$

Using the following partial derivatives with respect to c_r

$$\begin{aligned}
\frac{\partial \overline{\mathbf{h}}(\mathbf{c})}{\partial c_r} &= 2\overline{\mathbf{m}}_r \odot \left(\overline{\mathbf{M}}\mathbf{c} + \mathbf{m}_R\right) \\
\frac{\partial (\mathbf{y} - \overline{\mathbf{M}}\mathbf{c} - \mathbf{m}_R)}{\partial c_r} &= -\overline{\mathbf{m}}_r \\
\frac{\partial (\mathbf{y} - \overline{\mathbf{M}}\mathbf{c} - \mathbf{m}_R)^T \overline{\mathbf{h}}(\mathbf{c})}{\partial c_r} &= -\overline{\mathbf{m}}_r^T \overline{\mathbf{h}}(\mathbf{c}) + 2(\mathbf{y} - \mathbf{M}\boldsymbol{a})^T \left(\overline{\mathbf{m}}_r \odot \left(\overline{\mathbf{M}}\mathbf{c} + \mathbf{m}_R\right)\right) \\
\frac{\partial \overline{\mathbf{h}}(\mathbf{c})^T \overline{\mathbf{h}}(\mathbf{c})}{\partial c_r} &= 2\overline{\mathbf{h}}(\mathbf{c})^T \frac{\partial \overline{\mathbf{h}}(\mathbf{c})}{\partial c_r}
\end{aligned}$$

and the usual differentiation rules, we obtain

$$\begin{aligned}
\frac{\partial \check{b}^*(\mathbf{c})}{\partial c_r} &= \frac{1}{\left\|\overline{\mathbf{h}}(\mathbf{c})\right\|^4} \left\{ \frac{\partial (\mathbf{y} - \overline{\mathbf{M}}\mathbf{c} - \mathbf{m}_R)^T \overline{\mathbf{h}}(\mathbf{c})}{\partial c_r} \left[\overline{\mathbf{h}}(\mathbf{c})^T \overline{\mathbf{h}}(\mathbf{c})\right] \right. \\
&- \left. (\mathbf{y} - \overline{\mathbf{M}}\mathbf{c} - \mathbf{m}_R)^T \overline{\mathbf{h}}(\mathbf{c}) \frac{\partial \overline{\mathbf{h}}(\mathbf{c})^T \overline{\mathbf{h}}(\mathbf{c})}{\partial c_r} \right\}.
\end{aligned}$$

Finally, the partial derivative of the estimated spectrum $\check{\mathbf{y}}^*(\mathbf{c})$ with respect to c_r is

$$\frac{\partial \check{\mathbf{y}}^*(\mathbf{c})}{\partial c_r} = \overline{\mathbf{m}}_r + \frac{\partial \check{b}^*(\mathbf{c})}{\partial c_r} \overline{\mathbf{h}}(\mathbf{c}) + \check{b}^*(\mathbf{c}) \frac{\partial \overline{\mathbf{h}}(\mathbf{c})}{\partial c_r}.$$

Appendix C

Derivation of the potential functions associated the UPPNMM algorithm

C.1 Derivation of the potential function associated with $\boldsymbol{z}_n$

The potential energy (1.61) can be rewritten

$$U(\boldsymbol{z}_n) = U_1(\boldsymbol{a}_n) + U_2(\boldsymbol{z}_n) \tag{C.1}$$

where

$$\begin{aligned} U_1(\boldsymbol{a}_n) &= \frac{1}{2}\left[\mathbf{y}_n - \mathbf{g}_n\left(\mathbf{M}\boldsymbol{a}_n\right)\right]^T \boldsymbol{\Sigma}^{-1}\left[\mathbf{y}_n - \mathbf{g}_n\left(\mathbf{M}\boldsymbol{a}_n\right)\right], \\ U_2(\boldsymbol{z}_n) &= -\sum_{r=1}^{R-1} \log\left(z_{r,n}^{R-r-1}\right). \end{aligned}$$

Partial derivatives of $U(\boldsymbol{z}_n)$ with respect to $\boldsymbol{z}_n$ is obtained using the classical chain rule

$$\frac{\partial U(\boldsymbol{z}_n)}{\partial \boldsymbol{z}_n} = \frac{\partial U_1(\boldsymbol{a}_n)}{\partial \boldsymbol{a}_n}\frac{\partial \boldsymbol{a}_n}{\partial \boldsymbol{z}_n} + \frac{\partial U_2(\boldsymbol{z}_n)}{\partial \boldsymbol{z}_n}.$$

Straightforward computations lead to

$$\frac{\partial U_1(\boldsymbol{a}_n)}{\partial \boldsymbol{a}_n} = -\left[\mathbf{y}_n - \mathbf{g}_n\left(\mathbf{M}\boldsymbol{a}_n\right)\right]^T \boldsymbol{\Sigma}^{-1}\left[\mathbf{M} + 2b_n\left(\mathbf{M}\boldsymbol{a}_n\mathbf{1}_R^T\right)\odot\mathbf{M}\right]$$

$$\begin{aligned} \frac{\partial a_{r,n}}{\partial z_{i,n}} &= \begin{cases} 0 & \text{if} \quad i > r \\ \dfrac{a_{r,n}}{z_{i,n}-1} & \text{if} \quad i = r \\ \dfrac{a_{r,n}}{z_{i,n}} & \text{if} \quad i < r \end{cases} \\ \frac{\partial U_2(\boldsymbol{z}_n)}{\partial z_{i,n}} &= -\frac{R-i-1}{z_{i,n}}. \end{aligned} \tag{C.2}$$

C.2 Derivation of the potential functions associated with the endmember matrix

Similarly, the potential energy (1.63) can be rewritten

$$V(\mathbf{m}_{\ell,:}) = V_1(\mathbf{t}_\ell) + V_2(\boldsymbol{z}_n) \tag{C.3}$$

with $\mathbf{t}_\ell = \mathbf{A}^T\mathbf{m}_{\ell,:} + \text{diag}\left(\boldsymbol{b}\right)\left[\left(\mathbf{A}^T\mathbf{m}_{\ell,:}\right) \odot \left(\mathbf{A}^T\mathbf{m}_{\ell,:}\right)\right]$ and

$$\begin{aligned} V_1(\mathbf{t}_\ell) &= \frac{\|\mathbf{y}_{\ell,:} - \mathbf{t}_\ell\|^2}{2\sigma_\ell^2} \\ V_2(\mathbf{m}_{\ell,:}) &= \frac{\|\mathbf{m}_{\ell,:} - \bar{\mathbf{m}}_{\ell,:}\|^2}{2s^2}. \end{aligned}$$

The partial derivatives of the potential energy (1.63) can be obtained using the chain rule

$$\frac{\partial V(\mathbf{m}_{\ell,:})}{\partial \mathbf{m}_{\ell,:}} = \frac{\partial V_1(\mathbf{t}_\ell)}{\partial \mathbf{t}_\ell}\frac{\partial \mathbf{t}_\ell}{\partial \mathbf{m}_{\ell,:}} + \frac{\partial V_2(\mathbf{m}_{\ell,:})}{\partial \mathbf{m}_{\ell,:}}$$

and

$$\begin{aligned} \frac{\partial V_1(\mathbf{t}_\ell)}{\partial \mathbf{t}_\ell} &= -\frac{(\mathbf{y}_{\ell,:} - \mathbf{t}_\ell)^T}{\sigma_\ell^2} \\ \frac{\partial \mathbf{t}_\ell}{\partial \mathbf{m}_{\ell,:}} &= \mathbf{A}^T + 2\text{diag}\left(\boldsymbol{b}\right)\left[\left(\mathbf{A}^T\mathbf{m}_{\ell,:}\mathbf{1}_R^T\right) \odot \mathbf{A}^T\right] \\ \frac{\partial V_2(\mathbf{m}_{\ell,:})}{\partial \mathbf{m}_{\ell,:}} &= \frac{(\mathbf{m}_{\ell,:} - \bar{\mathbf{m}}_{\ell,:})^T}{s^2} \end{aligned}$$

Appendix D

On the linear mapping between latent variables and abundances

Consider a hyperspectral data set composed of R nonlinearly mixed endmembers. In the noise-free case, we assume that each observed pixel $\mathbf{y}_n$ can be expressed as

$$\mathbf{y} = \mathbf{W}_0 \boldsymbol{\psi}[\boldsymbol{a}], \tag{D.1}$$

where $\mathbf{W}_0$ an $L \times D$ matrix such that $\text{rank}(\mathbf{W}_0) = D$, $D = R(R+1)/2$ and $\boldsymbol{\psi}[\cdot]$ is defined by

$$\begin{aligned} \boldsymbol{\psi}: \quad \mathbb{R}^R &\to \mathbb{R}^D \\ \boldsymbol{a} &\mapsto \boldsymbol{\psi}[\boldsymbol{a}] = [a_1, \ldots, a_R, a_1 a_2 \ldots, a_{R-1} a_R]^T, \end{aligned} \tag{D.2}$$

as in (2.6). The $R \times 1$ vector $\boldsymbol{a}$ contains the abundances of the R components and satisfies the following positivity and sum-to one constraints

$$\sum_{r=1}^{R} a_r = 1, \quad a_r \geq 0, \forall r \in \{1, \ldots, R\}. \tag{D.3}$$

In particular, the abundance vectors belong to the following set

$$\mathcal{P} = \left\{ \boldsymbol{a} \middle| \sum_{r=1}^{R} a_r = 1 \right\} \tag{D.4}$$

We want to show the following statement

$$\left(\exists \mathbf{W} \in \mathcal{M}_{L,D}(\mathbb{R}), \forall \boldsymbol{a} \in \mathcal{P}, \exists \boldsymbol{x} \in \mathcal{P}, \mathbf{W}_0 \boldsymbol{\psi}[\boldsymbol{a}] = \mathbf{W} \boldsymbol{\psi}[\boldsymbol{x}]\right) \Rightarrow \left(\exists \mathbf{V} \in \mathcal{M}_{R,R}(\mathbb{R}), \forall \boldsymbol{a} \in \mathcal{P}, \boldsymbol{x} = \mathbf{V}\boldsymbol{a}\right). \tag{D.5}$$

In other words, if we can a matrix $\mathbf{W}$ such that $\mathbf{y} = \mathbf{W}\boldsymbol{\psi}[\boldsymbol{x}]$ for all the pixels of the observed image, there is necessarily a linear relation between the variables $\boldsymbol{x}$ and the abundances $\boldsymbol{a}$ of interest. This demonstration is decomposed into three main steps

1. First, we will show that $\dim(\text{span}(\mathbf{y})) = D$, where $F_{\mathbf{y}} = \text{span}(\mathbf{y})$.
2. Second, we will show that $\text{rank}(\mathbf{W}) = \text{rank}(\mathbf{W}_0) = D$.
3. Third, we will conclude that $\exists \mathbf{V} \in \mathcal{M}_{R,R}(\mathbb{R}), \forall \boldsymbol{a} \in \mathcal{P}, \boldsymbol{x} = \mathbf{V}\boldsymbol{a}$

D.1 Dimension of the subspace spanned by y

Consider the set of monomials

$$f_{\boldsymbol{a}} = \left\{ \left(1 - \sum_{r=1}^{R-1} a_r\right)^{\delta_R} \left[\prod_{r=1}^{R-1} a_r^{\delta_r}\right] \left| \forall r \in \{1, \dots R\}, \delta_r \in \{0,1\}, 1 \leq \sum_{r=1}^{R} \delta_r \leq 2 \right.\right\} \tag{D.6}$$

which consists of the elements of $\boldsymbol{\psi}[\boldsymbol{a}]$ when $\boldsymbol{a} \in \mathcal{P}$ and $F_{\boldsymbol{a}}$ the space spanned by the elements of $f_{\boldsymbol{a}}$, i.e.,

$$F_{\boldsymbol{a}} = \text{span}(f_{\boldsymbol{a}}). \tag{D.7}$$

We first want to show that $\dim(F_{\boldsymbol{a}}) = D$, i.e., that $f_{\boldsymbol{a}}$ is a basis of $F_{\boldsymbol{a}}$. Note that it is straightforward to see that $\dim(F_{\boldsymbol{a}}) \leq \text{card}(f_{\boldsymbol{a}}) = D$.

Consider the set

$$g^*_{\boldsymbol{a}} = \left\{ \prod_{r=1}^{R-1} a_r^{\delta_r} \left| \forall r \in \{1, \dots R-1\}, \delta_r \in \{0,1,2,3,4\}, 0 \leq \sum_{r=1}^{R-1} \delta_r \leq 4 \right.\right\} \tag{D.8}$$

which is a basis of the $(R-1)$-variate polynomials (in $\boldsymbol{a}_{\setminus R}$) such that $\sum_{r=1}^{R-1} \delta_r \leq 4$. In other words, the elements of $g^*_{\boldsymbol{a}}$ are linearly independent. In particular, the elements of

$$g_{\boldsymbol{a}} = \left\{ \prod_{r=1}^{R-1} a_r^{\delta_r} \left| \forall r \in \{1, \dots R-1\}, \delta_r \in \{0,1,2\}, 0 \leq \sum_{r=1}^{R-1} \delta_r \leq 2 \right.\right\} \tag{D.9}$$

are linearly independent, and thus form a basis of the subspace $G_{\boldsymbol{a}} = \text{span}(g_{\boldsymbol{a}})$ which consists of the $(R-1)$-variate polynomials such that $\sum_{r=1}^{R-1} \delta_r \leq 2$. It can be easily shown that $\dim(G_{\boldsymbol{a}}) = D$. To prove that $f_{\boldsymbol{a}}$ is a basis of $F_{\boldsymbol{a}}$, we will show that $F_{\boldsymbol{a}} = G_{\boldsymbol{a}}$. We first prove that each basis element of $g_{\boldsymbol{a}}$ belongs to $F_{\boldsymbol{a}}$.

$$\begin{aligned} 1 &= \prod_{r=1}^{R-1} a_r^0 \in G_{\boldsymbol{a}} \\ &= \left(1 - \sum_{r=1}^{R-1} a_r\right)^0 \left[\prod_{r=1}^{R-1} a_r^0\right] \in F_{\boldsymbol{a}}. \end{aligned} \tag{D.10}$$

$$\begin{aligned} a_r &= a_r^1 \prod_{i=1, i\neq r}^{R-1} a_i^0 \in G_{\boldsymbol{a}} \\ &= \left(1 - \sum_{r=1}^{R-1} a_r\right)^0 a_r^1 \left[\prod_{i=1, i\neq r}^{R-1} a_i^0\right] \in F_{\boldsymbol{a}}, \quad \forall r \in \{1, \dots, R-1\}. \end{aligned} \tag{D.11}$$

$$\begin{aligned} a_i a_j &= a_i^1 a_j^1 \prod_{r=1, r\neq i, r\neq j}^{R-1} a_r^0 \in G_{\boldsymbol{a}} \\ &= \left(1 - \sum_{r=1}^{R-1} a_r\right)^0 a_i^1 a_j^1 \left[\prod_{r=1, r\neq i, r\neq j}^{R-1} a_r^0\right] \in F_{\boldsymbol{a}}, \quad \forall i, j \in \{1, \dots R-1\}, i \neq j. \end{aligned} \tag{D.12}$$

$$\begin{aligned} a_r^2 &= a_r^2 \prod_{i=1, i\neq r}^{R-1} a_i^0 \in G_{\boldsymbol{a}} \\ &= a_r - a_r \left(1 - \sum_{i=1}^{R-1} a_i\right) - \sum_{i=1, i\neq r}^{R-1} a_r a_i \in F_{\boldsymbol{a}}, \quad \forall r \in \{1, \dots, R-1\}. \end{aligned} \tag{D.13}$$

Finally, we obtain $g_{\boldsymbol{a}} \subset F_{\boldsymbol{a}}$, i.e., $G_{\boldsymbol{a}} \subset F_{\boldsymbol{a}}$ and

$$D = \dim(G_{\boldsymbol{a}}) \leq \dim(F_{\boldsymbol{a}}) \leq D \tag{D.14}$$

which leads to $\dim(F_{\boldsymbol{a}}) = D$ and shows that $f_{\boldsymbol{a}}$ is a basis of $F_{\boldsymbol{a}}$. Since $\text{rank}(\mathbf{W}_0) = D$ and $\dim(F_{\boldsymbol{a}}) = D$, we obtain $\dim(F_{\mathbf{y}}) = D$ where $F_{\mathbf{y}} = \text{span}(\mathbf{y})$

D.2 Existence and rank of W

The existence of the matrix $\mathbf{W}$ in (D.5) is straightforward. For $\mathbf{W} = \mathbf{W}_0$, $\boldsymbol{x} = \mathbf{I}_R \boldsymbol{a}$ is a solution of $\mathbf{W}_0 \boldsymbol{\psi}[\boldsymbol{a}] = \mathbf{W}\boldsymbol{\psi}[\boldsymbol{x}]$. In the previous paragraph, we have shown that $\dim(F_{\mathbf{y}}) = D$ if $\text{rank}(\mathbf{W}_0) = D$ since $\dim(F_{\boldsymbol{a}}) = D$. Let $\mathbf{W}$ be an $L \times D$ matrix such that

$$\forall \boldsymbol{a} \in \mathcal{P}, \exists \boldsymbol{x} \in \mathcal{P}, \mathbf{W}_0 \boldsymbol{\psi}[\boldsymbol{a}] = \mathbf{W}\boldsymbol{\psi}[\boldsymbol{x}]. \tag{D.15}$$

Using the results of the previous paragraph, we obtain $\dim(F_{\boldsymbol{x}}) = D$, where $F_{\boldsymbol{x}} = \text{span}(f_{\boldsymbol{x}})$ and

$$f_{\boldsymbol{x}} = \left\{ \left(1 - \sum_{r=1}^{R-1} xr\right)^{\delta_R} \left[\prod_{r=1}^{R-1} x_r^{\delta_r}\right] \middle| \forall r \in \{1, \dots R\}, \delta_r \in \{0,1\}, 1 \leq \sum_{r=1}^{R} \delta_r \leq 2 \right\}. \tag{D.16}$$

Thus, since $\dim(F_{\mathbf{y}}) = D$, we obtain $\text{rank}(\mathbf{W}) = \text{rank}(\mathbf{W}_0) = D$ and $F_{\mathbf{y}} = \text{span}(\mathbf{W}) = \text{span}(\mathbf{W}_0)$.

D.3 Relation between $\boldsymbol{a}$ and $\boldsymbol{x}$

The columns of $\mathbf{W}_0$ and $\mathbf{W}$ span the same subspace ($F_{\mathbf{y}}$). Consequently, there exists an $D \times D$ invertible matrix $\mathbf{T} \in \mathcal{GL}_D(\mathbb{R})$ such that $\mathbf{W}_0 = \mathbf{W}\mathbf{T}$, i.e.,

$$\exists \mathbf{T} \in \mathcal{GL}_D(\mathbb{R}) \forall \boldsymbol{a} \in \mathcal{P}, \exists \boldsymbol{x} \in \mathcal{P}, \boldsymbol{\psi}[\boldsymbol{a}] = \mathbf{T}^{-1} \boldsymbol{\psi}[\boldsymbol{x}], \tag{D.17}$$

which means that $F_{\boldsymbol{a}} = F_{\boldsymbol{x}}$. The aim of this paragraph is to find all the basis of $F_{\boldsymbol{a}}$ that can be expressed as

$$f_{\boldsymbol{x}} = \left\{ \left(1 - \sum_{r=1}^{R-1} xr\right)^{\delta_R} \left[\prod_{r=1}^{R-1} x_r^{\delta_r}\right] \middle| \forall r \in \{1, \dots R\}, \delta_r \in \{0,1\}, 1 \leq \sum_{r=1}^{R} \delta_r \leq 2 \right\} \tag{D.18}$$

where $\boldsymbol{x} \in \mathcal{P}$.

Consider $\boldsymbol{a} \in \mathcal{P}$ and $\boldsymbol{x} = [x_1, \dots, x_R] \in \mathcal{P}$ such that $\text{span}(\boldsymbol{\psi}_{\boldsymbol{x}}) = F_{\boldsymbol{a}}$. For $r = 1, \dots, (R-1)$, $x_r \in F_{\boldsymbol{a}} = G_{\boldsymbol{a}}$. Consequently, x_r can be decomposed onto the basis $g_{\boldsymbol{a}}$ as follows

$$x_r = \sum_{d=1}^{D} \gamma_{r,d} \left[\prod_{i=1}^{R-1} a_i^{\delta_{i,d}^{(r)}}\right], \quad r = 1, \dots, (R-1) \tag{D.19}$$

where

$$\begin{aligned} \forall d = 1, \dots, D, \forall i = 1, \dots, (R-1), 0 \leq \delta_{i,d}^{(d)} \leq 2 \\ \forall d = 1, \dots, D, 0 \leq \sum_{i=1}^{R-1} \delta_{i,d}^{(r)} \leq 2 \end{aligned} \tag{D.20}$$

and $\{\gamma_{r,d}\}_{d=1,\ldots,D}$ are the coefficients of the decomposition of x_r onto $g_{\boldsymbol{a}}$. Moreover, for $r, r'(r \neq r')$

$$x_r x_{r'} = \left(\sum_{d=1}^{D} \gamma_{r,d} \left[\prod_{i=1}^{R-1} a_i^{\delta_{i,d}^{(r)}}\right]\right)\left(\sum_{d=1}^{D} \gamma_{r',d} \left[\prod_{i=1}^{R-1} a_i^{\delta_{i,d}^{(r')}}\right]\right) \in \text{span}(g_{\boldsymbol{a}}^{+}). \tag{D.21}$$

However, if $f_{Vlatents}$ is a basis of $F_{\boldsymbol{a}}$, then $x_r x_{r'} \in F_{\boldsymbol{a}}$, i.e.,

$$x_r x_{r'} = \sum_{d=1}^{D} \gamma_{(r,r'),d} \left[\prod_{i=1}^{R-1} a_i^{\delta_{i,d}^{(r,r')}}\right] \tag{D.22}$$

where $\left\{\gamma_{(r,r'),d}\right\}_{d=1,\ldots,D}$ are the coefficients of the decomposition of $x_r x_{r'}$ onto $g_{\boldsymbol{a}}$.
Assume there exists $r_0 \in 1, \ldots, (R-1)$ and $d_0 \in 1, \ldots, D$ such that

$$x_{r_0} = \sum_{d=1}^{D} \gamma_{r_0,d} \left[\prod_{i=1}^{R-1} a_i^{\delta_{i,d}^{(r_0)}}\right]$$
$$\sum_{i=1}^{R-1} \delta_{i,d_0}^{(r_0)} = 2. \tag{D.23}$$

Then, for $r \neq r_0$, x_r must be a constant to ensure that $x_{r_0} x_r \in F_{\boldsymbol{a}}$, which is impossible since 1 already belongs to $F_{\boldsymbol{a}}$. Consequently,

$$\begin{aligned} x_r &= \gamma_{r,0} + \sum_{i=1}^{R-1} \gamma_{r,i} a_i \\ &= \gamma_{r,0} \sum_{r=1}^{R} a_r + \sum_{i=1}^{R-1} \gamma_{r,i} a_i \\ &= \sum_{i=1}^{R-1} (\gamma_{r,i} + \gamma_{r,0}) a_i + \gamma_{r,0} a_R \\ &= \mathbf{v}_r^T \boldsymbol{a} \quad r = 1, \ldots, (R-1) \end{aligned} \tag{D.24}$$

where $\mathbf{v}_r = [\gamma_{r,1} + \gamma_{r,0}, \ldots, \gamma_{r,R-1} + \gamma_{r,0}, \gamma_{r,0}]^T$ is an $(R-1) \times 1$ vector. Straightforward computations lead to

$$\boldsymbol{x} = \mathbf{V}\boldsymbol{a}, \tag{D.25}$$

where

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}^{-} \\ \mathbf{1}_R^T - \mathbf{1}_{R-1}\mathbf{V}^{-} \end{bmatrix} \tag{D.26}$$

and $\mathbf{V}^{-} = [\mathbf{v}_1, \ldots, \mathbf{v}_{R-1}]^T$ is an $(R-1) \times R$ matrix.

D.4 Scenario where $\text{rank}(\mathbf{W}_0) < D$

In the previous paragraphs, we have shown that if $\text{rank}(\mathbf{W}_0) = D$, then $F_{\mathbf{y}} = \text{span}(\mathbf{W}) = \text{span}(\mathbf{W}_0)$. Conversely, if $\text{rank}(\mathbf{W}_0) < D$, then $F_{\mathbf{y}} = \text{span}(\mathbf{W}_0) \subset \text{span}(\mathbf{W})$, i.e., there still exists a matrix $\mathbf{T}$ such that $\mathbf{W}_0 = \mathbf{W}\mathbf{T}$ but the matrix $\mathbf{T}$ is not necessarily invertible. In that case, the latent variables are not necessarily linear transformations

of the abundances of interest. For instance, consider the scenario of $R = 3$ linear mixed pixels, i.e., $\mathbf{W}_0 = [\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3, \mathbf{0}_{L,3}]$ ($D = 6$). If we set

$$\begin{aligned} a_1 &= x_1 + x_1^2 \\ a_2 &= x_2 - x_1^2 \\ a_3 &= x_3, \end{aligned} \tag{D.27}$$

we obtain

$$\begin{aligned} \mathbf{y} &= \sum_{r=1}^{3} \mathbf{m}_r a_r \\ &= \mathbf{m}_1 x_1 + \mathbf{m}_2 x_2 + \mathbf{m}_3 x_3 + (\mathbf{m}_1 - \mathbf{m}_2) x_1^2 \\ &= \mathbf{m}_1 x_1 + \mathbf{m}_2 x_2 + \mathbf{m}_3 x_3 + (\mathbf{m}_1 - \mathbf{m}_2)(x_1 - x_2 x_3 - x_1 x_3) \\ &= (2\mathbf{m}_1 - \mathbf{m}_2) x_1 + \mathbf{m}_2 x_2 + \mathbf{m}_3 x_3 - (\mathbf{m}_1 - \mathbf{m}_2) x_2 x_3 - (\mathbf{m}_1 - \mathbf{m}_2) x_1 x_3 \\ &= \mathbf{W} \boldsymbol{\psi}[\boldsymbol{x}]. \end{aligned} \tag{D.28}$$

where $\mathbf{W} = [2\mathbf{m}_1 - \mathbf{m}_2, \mathbf{m}_2, \mathbf{m}_3, \mathbf{0}_L, \mathbf{m}_2 - \mathbf{m}_1 \mathbf{m}_2 - \mathbf{m}_1]$ Finally, $(\mathbf{W}, \boldsymbol{x})$ where $\boldsymbol{x}$ is defined in (D.27) is a solution of $\mathbf{W}_0 \boldsymbol{\psi}[\boldsymbol{a}] = \mathbf{W} \boldsymbol{\psi}[\boldsymbol{x}]$ but the latent variables in $\boldsymbol{x}$ are not linear combinations of the abundances of interest. In practice, we want to apply nonlinear unmixing procedure to hyperspectral images where nonlinearities occur. Consequently the rank of the matrix $\mathbf{W}_0$ will often be close to D. However, if $\mathbf{W}_0$ is ill-conditioned, the prior distribution for the latent variables, which enforces spectrally close pixels to have close latent variables in (2.19), will also enforce linear relations between abundances and latent variables.

Appendix E

Structured covariance matrices of Gaussian prior for correlated Gaussian processes

Eq. (2.10) can be rewritten

$$\mathbf{y}_{\ell,:} \quad = \quad \mathbf{\Psi}_x \mathbf{w}_\ell + \mathbf{e}_{\ell,:}, \quad \ell = 1, \ldots, L \tag{E.1}$$

where $\mathbf{\Psi}_x = [\boldsymbol{\psi}_x(1), \ldots, \boldsymbol{\psi}_x(N)]^T$ is an $N \times D$ matrix and $\mathbf{e}_{\ell,:}$ is an $N \times 1$ noise vector such that $\mathbf{e}_{\ell,:} \sim \mathcal{N}\left(\mathbf{0}_N, \sigma^2 \mathbf{I}_N\right)$ yielding

$$\tilde{\mathbf{y}} \quad = \quad (\mathbf{I}_L \otimes \mathbf{\Psi}_x)\tilde{\mathbf{w}} + \tilde{\mathbf{e}} \tag{E.2}$$

where $\tilde{\mathbf{w}} = [\mathbf{w}_1^T, \ldots, \mathbf{w}_L^T]^T$ is an $LD \times 1$ vector, $\tilde{\mathbf{y}} = [\mathbf{y}_{1,:}^T, \ldots, \mathbf{y}_{L,:}^T]^T$ and $\tilde{\mathbf{e}} = [\mathbf{e}_{1,:}{}^T, \ldots, \mathbf{e}_{L,:}{}^T]^T$ are $NL \times 1$ vectors, $\tilde{\mathbf{e}} \sim \mathcal{N}\left(\mathbf{0}_{NL}, \sigma^2 \mathbf{I}_{NL}\right)$ and $\otimes$ denotes the Kronecker product. For ease of marginalization of (2.13), a Gaussian prior is usually assigned for $\tilde{\mathbf{w}}$ (Lawrence, 2003). Using centered data, it makes sense to assign a zero-mean prior for $\tilde{\mathbf{w}}$. Introducing correlation through structured covariance matrix yields the following Gaussian prior for $\tilde{\mathbf{w}}$

$$\tilde{\mathbf{w}} \sim \mathcal{N}\left(\mathbf{0}_{LD}, \mathbf{\Phi} \otimes \mathbf{\Gamma}\right) \tag{E.3}$$

where $\mathbf{\Phi}$ (resp. $\mathbf{\Gamma}$) is an $L \times L$ (resp. a $D \times D$) covariance matrix that reflects the correlation between the rows (resp. the columns) of $\mathbf{W}$. More precisely, the covariance matrix $\mathbf{\Gamma}$ reflects the prior correlations between the spectra whereas $\mathbf{\Phi}$ reflects prior correlations between the L spectral bands. It makes sense to consider correlations between spectra since the interaction spectrum between the components i and j is likely to be related to the spectra of the components i and j. Consequently, $\mathbf{\Gamma}$ is assumed to be a non-diagonal matrix. Similarly, since the reflectances observed at spectrally close bands are highly related, it seems reasonable to assume that $\mathbf{\Psi}$ is not diagonal. Let assume that $\mathbf{\Gamma}$ and $\mathbf{\Phi}$ are symmetric matrices with no particular block structure. Marginalizing out $\tilde{\mathbf{w}}$ leads to the following marginalized likelihood

$$\begin{aligned} f(\tilde{\mathbf{y}}|\mathbf{\Psi}_x, \mathbf{\Gamma}, \mathbf{\Phi}, \sigma^2) \quad &= \quad \int f(\tilde{\mathbf{y}}|\tilde{\mathbf{w}}, \mathbf{\Psi}_x, \mathbf{\Gamma}, \mathbf{\Phi}, \sigma^2) f\left(\tilde{\mathbf{w}}\right) \mathrm{d}\tilde{\mathbf{w}} \\ &\propto \quad |\mathbf{\Upsilon}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\tilde{\mathbf{y}}^T \mathbf{\Upsilon}^{-1} \tilde{\mathbf{y}}\right] \end{aligned} \tag{E.4}$$

where $\mathbf{\Upsilon} = \mathbf{\Phi} \otimes (\mathbf{\Psi}_x \mathbf{\Gamma} \mathbf{\Psi}_x^T) + \sigma^2 \mathbf{I}_{NL}$ is an $NL \times NL$ covariance matrix.

Appendix F

Partial derivatives of the log-posterior distribution of the LL-GPLVM

The joint log-posterior distribution is given by (up to an additive constant)

$$\mathcal{L}_p(\mathbf{U}, \mathbf{X}_{\setminus R}, \sigma^2, s^2) = \mathcal{L} - \frac{\gamma}{2}\text{tr}\left[\left((\mathbf{I}_N - \mathbf{\Lambda}_{\text{LLE}})\mathbf{X}_{\setminus R}^T\mathbf{B}\right)\left((\mathbf{I}_N - \mathbf{\Lambda}_{\text{LLE}})\mathbf{X}_{\setminus R}^T\mathbf{B}\right)^T\right]$$

on the admissible set for $\mathbf{U}, \mathbf{X}_{\setminus R}, \sigma^2, s^2$, where $\mathbf{B} = [\mathbf{I}_{R-1}, \mathbf{1}_{R-1}]$ is an $(R-1) \times R$ matrix and the log-likelihood $\mathcal{L}$ is

$$\mathcal{L} = -\frac{NL}{2}\ln(2\pi) - \frac{L}{2}\ln(|\mathbf{\Sigma}|) - \frac{1}{2}\text{trace}(\mathbf{\Sigma}^{-1}\bar{\mathbf{Y}}^T\bar{\mathbf{Y}})$$

with $\bar{\mathbf{Y}} = \mathbf{Y} - \bar{\mathbf{P}}\mathbf{C}^T$ and $\mathbf{\Sigma} = s^2\mathbf{C}\mathbf{C}^T + \sigma^2\mathbf{I}_N$. The partial derivatives of $\mathcal{L}$ with respect to the covariance matrix $\mathbf{\Sigma}$ is given by

$$\frac{\partial\mathcal{L}}{\partial\mathbf{\Sigma}} = -\frac{L}{2}\mathbf{\Sigma}^{-1} + \frac{1}{2}\mathbf{\Sigma}^{-1}\bar{\mathbf{Y}}^T\bar{\mathbf{Y}}\mathbf{\Sigma}^{-1}.$$

The partial derivative of $\mathcal{L}$ with respect to the covariance matrix $\bar{\mathbf{Y}}$ is given by

$$\frac{\partial\mathcal{L}}{\partial\bar{\mathbf{Y}}} = -\mathbf{\Sigma}^{-1}\bar{\mathbf{Y}}^T.$$

Using the classical chain rules, we obtain

$$\begin{aligned}
\frac{\partial\mathcal{L}}{\partial s^2} &= \text{trace}\left(\frac{\partial\mathcal{L}}{\partial\mathbf{\Sigma}}\mathbf{C}\mathbf{C}^T\right) \\
\frac{\partial\mathcal{L}}{\partial\sigma^2} &= \text{trace}\left(\frac{\partial\mathcal{L}}{\partial\mathbf{\Sigma}}\right) \\
\frac{\partial\mathcal{L}}{\partial\mathbf{C}^T} &= 2s^2\left(\frac{\partial\mathcal{L}}{\partial\mathbf{\Sigma}}\right)\mathbf{C} + \mathbf{\Sigma}^{-1}\bar{\mathbf{Y}}^T\bar{\mathbf{P}}
\end{aligned}$$

which leads to

$$\begin{aligned}
\frac{\partial\mathcal{L}}{\partial\mathbf{\Psi}_x^T} &= 2s^2\left(\frac{\partial\mathcal{L}}{\partial\mathbf{\Sigma}}\right)\mathbf{C}\mathbf{U}^T + \mathbf{\Sigma}^{-1}\bar{\mathbf{Y}}^T\bar{\mathbf{P}}\mathbf{U}^T \\
\frac{\partial\mathcal{L}}{\partial\mathbf{U}^T} &= 2s^2\mathbf{\Psi}_x^T\left(\frac{\partial\mathcal{L}}{\partial\mathbf{\Sigma}}\right)\mathbf{C} + \mathbf{\Psi}_x^T\mathbf{\Sigma}^{-1}\bar{\mathbf{Y}}^T\bar{\mathbf{P}}
\end{aligned}$$

The partial derivatives of the log-likelihood w.r.t. to $\mathbf{X}_{\setminus R}$ are obtained using

$$\frac{\partial \mathcal{L}}{\partial x_{r,n}} = \left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{\psi}_x(n)}\right)^T \frac{\partial \boldsymbol{\psi}_x(n)}{\partial x_{r,n}}, \quad r = 1, \ldots, R-1, n = 1, \ldots, N.$$

Finally, the partial derivatives of the log-posterior w.r.t. to $\mathbf{X}_{\setminus R}$ are given by

$$\frac{\partial \mathcal{L}_p}{\partial \mathbf{X}_{\setminus R}{}^T} = \frac{\partial \mathcal{L}}{\partial \mathbf{X}_{\setminus R}{}^T} - \gamma \left(\mathbf{I}_N - \boldsymbol{\Lambda}_{\text{LLE}}^T\right) \left(\mathbf{I}_N - \boldsymbol{\Lambda}_{\text{LLE}}\right)^T \mathbf{X}_{\setminus R}^T \mathbf{B}\mathbf{B}^T.$$

Appendix G

Fisher information matrix associated with the supervised unmixing problem using the PPNMM

The likelihood function of $\mathbf{y}$ can be expressed as

$$f(\mathbf{y}|\boldsymbol{a}, b, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{L}{2}} \exp\left(-\frac{\|\mathbf{y} - \mathbf{g}(\mathbf{M}\boldsymbol{a})\|^2}{2\sigma^2}\right)$$

where $\mathbf{g}(\mathbf{M}\boldsymbol{a}) = \mathbf{M}\boldsymbol{a} + b(\mathbf{M}\boldsymbol{a}) \odot (\mathbf{M}\boldsymbol{a})$. The corresponding log-likelihood P can be written

$$P = \ln f(\mathbf{y}|\boldsymbol{a}, b, \sigma^2) = -\frac{L}{2}\ln(2\pi\sigma^2) - \frac{\|\mathbf{y} - \mathbf{g}(\mathbf{M}\boldsymbol{a})\|^2}{2\sigma^2}.$$

The partial derivatives of P with respect to the model parameters are

$$\begin{aligned}
\frac{\partial P}{\partial a_r} &= \frac{1}{2\sigma^2}\left[\mathbf{y} - \mathbf{g}(\mathbf{M}\boldsymbol{a})\right]^T \frac{\partial \mathbf{g}(\mathbf{M}\boldsymbol{a})}{\partial a_r} \\
\frac{\partial P}{\partial b} &= \frac{1}{2\sigma^2}\left[\mathbf{y} - \mathbf{g}(\mathbf{M}\boldsymbol{a})\right]^T \frac{\partial \mathbf{g}(\mathbf{M}\boldsymbol{a})}{\partial b} \\
\frac{\partial P}{\sigma^2} &= -\frac{L}{2\sigma^2} + \frac{\|\mathbf{y} - \mathbf{g}(\mathbf{M}\boldsymbol{a})\|^2}{2\sigma^4}.
\end{aligned}$$

Straightforward computations lead to

$$\begin{aligned}
-\mathrm{E}\left[\frac{\partial^2 P}{\partial a_i \partial a_j}\right] &= -\mathrm{E}\left[\frac{\partial^2 P}{\partial a_j \partial a_i}\right] \\
&= \frac{1}{2\sigma^2}\left(\frac{\partial \mathbf{g}(\mathbf{M}\boldsymbol{a})}{\partial a_i}\right)^T \frac{\partial \mathbf{g}(\mathbf{M}\boldsymbol{a})}{\partial a_j} \\
-\mathrm{E}\left[\frac{\partial^2 P}{\partial a_i \partial b}\right] &= -\mathrm{E}\left[\frac{\partial^2 P}{\partial b \partial a_i}\right] \\
&= \frac{1}{2\sigma^2}\left(\frac{\partial \mathbf{g}(\mathbf{M}\boldsymbol{a})}{\partial a_i}\right)^T \frac{\partial \mathbf{g}(\mathbf{M}\boldsymbol{a})}{\partial b} \\
-\mathrm{E}\left[\frac{\partial^2 P}{\partial^2 b}\right] &= \frac{1}{2\sigma^2}\left(\frac{\partial \mathbf{g}(\mathbf{M}\boldsymbol{a})}{\partial b}\right)^T \frac{\partial \mathbf{g}(\mathbf{M}\boldsymbol{a})}{\partial b}
\end{aligned}$$

where

$$\frac{\partial \mathbf{g}\left(\mathbf{M}\boldsymbol{a}\right)}{\partial a_r} = \mathbf{m}_r + 2b(\mathbf{M}\boldsymbol{a}) \odot \mathbf{m}_r$$
$$\frac{\partial \mathbf{g}\left(\mathbf{M}\boldsymbol{a}\right)}{\partial b} = (\mathbf{M}\boldsymbol{a}) \odot (\mathbf{M}\boldsymbol{a})$$

It can be easily shown that

$$\begin{aligned} -\mathrm{E}\left[\frac{\partial^2 P}{\partial a_i \partial \sigma^2}\right] &= -\mathrm{E}\left[\frac{\partial^2 P}{\partial \sigma^2 \partial a_i}\right] = 0 \\ -\mathrm{E}\left[\frac{\partial^2 P}{\partial b \partial \sigma^2}\right] &= -\mathrm{E}\left[\frac{\partial^2 P}{\partial \sigma^2 \partial b}\right] = 0 \\ -\mathrm{E}\left[\frac{\partial^2 P}{\partial^2 \sigma^2}\right] &= \frac{L}{2\sigma^4}. \end{aligned}$$

Appendix H

Alternative interpretation of the proposed RCA model

We consider a set of N observed pixel spectra $\mathbf{y}_n = [y_{n,1}, \ldots, y_{n,L}]^T, n \in \{1, \ldots, N\}$ where L is the number of spectral bands. Each of these spectra is defined as a linear combination of R known endmembers $\mathbf{m}_r$, contaminated by additional polynomials of the endmember spectra and additive noise

$$\begin{aligned} \mathbf{y}_n &= \sum_{r=1}^{R} a_{r,n}\mathbf{m}_r + \sum_{r=1}^{R} \gamma_{r,r,n}\mathbf{m}_r \odot \mathbf{m}_r + \sum_{i=1}^{R-1}\sum_{j=i+1}^{R} \gamma_{i,j,n}\sqrt{2}\mathbf{m}_i \odot \mathbf{m}_j + \mathbf{e}_n \\ &= \mathbf{M}\boldsymbol{a}_n + \mathbf{Q}\boldsymbol{\gamma}_n + \mathbf{e}_n, \quad n = 1, \ldots, N \end{aligned} \tag{H.1}$$

where $\mathbf{m}_r$ is the spectrum of the rth material present in the scene, $a_{r,n}$ is its corresponding proportion in the nth pixel and $\mathbf{e}_n$ is an additive independently and non identically distributed zero-mean Gaussian noise sequence with diagonal covariance matrix $\boldsymbol{\Sigma}_0 = \text{diag}\left(\boldsymbol{\sigma}^2\right)$, denoted as $\mathbf{e}_n \sim \mathcal{N}\left(\mathbf{0}_L, \boldsymbol{\Sigma}_0\right)$, where $\boldsymbol{\sigma}^2 = [\sigma_1^2, \ldots, \sigma_L^2]^T$. The parameters $\gamma_{r,r,n}$ control the amplitude of the nonlinear terms $\mathbf{m}_r \odot \mathbf{m}_r$ that model the double reflections involving only one material and the parameters $\gamma_{i,j,n}$ tune the amplitudes of the interactions between different endmembers. These parameters are gathered in the vector $\boldsymbol{\gamma}_n = [\gamma_{1,1,n}, \ldots, \gamma_{R,R,n}, \gamma_{1,2,n}, \ldots, \gamma_{R-1,R,n}]^T$ of size $R(R+1)/2 \times 1$. Moreover, $\mathbf{Q} = [\mathbf{m}_1 \odot \mathbf{m}_1, \ldots, \mathbf{m}_R \odot \mathbf{m}_R, \sqrt{2}\mathbf{m}_1 \odot \mathbf{m}_2, \ldots, \sqrt{2}\mathbf{m}_{R-1} \odot \mathbf{m}_R]$ is an $L \times R(R+1)/2$ matrix (as defined below (4.5.1)). The model (H.1) reduces to the bilinear models (Fan et al., 2009; Halimi et al., 2011a; Nascimento and Bioucas-Dias, 2009) when the coefficients $\gamma_{r,r,n}$ are all null. Moreover, it is similar to the model introduced by Meganem et al. (2013). Assume that the endmembers are known. The nonlinear SU problem consists then of estimating the abundance vectors $\{\boldsymbol{a}_n\}_{n=1,\ldots,N}$ when the nonlinearity parameters vectors $\{\boldsymbol{\gamma}_n\}_{n=1,\ldots,N}$ are unknown. Note that the number of unknown parameters can be very large, especially when the number of endmembers R is large. It has been previously mentioned that the LMM is sufficient to model accurately most of the observed pixels. Consequently, most of the nonlinearity parameters are expected to be small (or even null). It is also interesting to mention that the columns of $\mathbf{Q}$ are highly correlated which complicates the estimation of the nonlinearity parameters $\{\boldsymbol{\gamma}_n\}_{n=1,\ldots,N}$ (large estimation variance). To avoid estimating the large set of nonlinearity parameters, a simple solution consists of assigning a prior distribution to these nuisance parameters and of marginalizing them from the likelihood defined by the observation equation (H.1). Based on the physical considerations mentioned above, it makes sense to assign the following zero-mean Gaussian prior

$$\boldsymbol{\gamma}_n | z_n = k, s_k^2 \sim \mathcal{N}\left(\mathbf{0}_D, s_k^2 \mathbf{I}_D\right) \tag{H.2}$$

to the nonlinearity parameter vector associated with the nth pixel, where $z_n \in 1, \ldots, K$ is a label associated with the nth pixel (K is a given number of classes) and $D = R(R+1)/2$. The hyperparameter $s_k^2, k = 1, \ldots, K$ controls the variance of the nonlinearity parameters associated with the pixels that belong to the kth class. The K classes are characterized by different levels of nonlinearity. The main motivation for assigning Gaussian priors for the nonlinearity parameter vectors is that it allows these nuisance to be marginalized, as will be shown in the sequel. Using (H.1) and the noise statistical properties, the likelihood of the nth observed pixel $\mathbf{y}_n$ can be expressed as follows

$$f(\mathbf{y}_n|\mathbf{M}, \boldsymbol{a}_n, \boldsymbol{\gamma}_n, \boldsymbol{\sigma}^2) \propto \quad |\boldsymbol{\Sigma}_0|^{-1/2} \exp\left[-\frac{(\mathbf{y}_n - \mathbf{M}\boldsymbol{a}_n - \mathbf{Q}\boldsymbol{\gamma}_n)^T \boldsymbol{\Sigma}_0^{-1} (\mathbf{y}_n - \mathbf{M}\boldsymbol{a}_n - \mathbf{Q}\boldsymbol{\gamma}_n)}{2}\right]. \tag{H.3}$$

Assume now that the nth pixel belongs to the kth class ($k = 1, \ldots, K$). Marginalizing $\boldsymbol{\gamma}_n$ consists of computing

$$f(\mathbf{y}_n|\mathbf{M}, \boldsymbol{a}_n, \boldsymbol{\sigma}^2, z_n = k, s_k^2) = \int f(\mathbf{y}_n|\mathbf{M}, \boldsymbol{a}_n, \boldsymbol{\gamma}_n, \boldsymbol{\sigma}^2) f(\boldsymbol{\gamma}_n|z_n = k, s_k^2) \mathrm{d}\boldsymbol{\gamma}_n, \tag{H.4}$$

where $f(\boldsymbol{\gamma}_n|z_n = k, s_k^2)$ is defined in (H.2). In our case, the closed-form expression of this can be computed (Bishop, 1995, App. B), leading to

$$\mathbf{y}_n|\mathbf{M}, \boldsymbol{a}_n, \boldsymbol{\sigma}^2, z_n = k, s_k^2 \sim \mathcal{N}\left(\mathbf{M}\boldsymbol{a}_n, s_k^2 \mathbf{Q}\mathbf{Q}^T + \boldsymbol{\Sigma}_0\right). \tag{H.5}$$

Consider an additional class ($k = 0$) associated with linearly mixed pixels. If the nth pixels is linearly mixed (i.e., $z_n = 0$), Eq. (H.1) reduces to the classical LMM yielding

$$\mathbf{y}_n|\mathbf{M}, \boldsymbol{a}_n, \boldsymbol{\sigma}^2, z_n = 0 \sim \mathcal{N}\left(\mathbf{M}\boldsymbol{a}_n, \boldsymbol{\Sigma}_0\right). \tag{H.6}$$

Assuming independence between the observed pixels, the joint likelihood of the observation matrix $\mathbf{Y}$ can be expressed as

$$\begin{aligned} f(\mathbf{Y}|\mathbf{M}, \mathbf{A}, \mathbf{z}, \mathbf{s}, \boldsymbol{\sigma}^2) &= \prod_{n=1}^{N} f(\mathbf{y}_n|\mathbf{M}, \boldsymbol{a}_n, \boldsymbol{\sigma}^2, z_n, \mathbf{s}) \\ &\propto \prod_{k=0}^{K-1} \prod_{n \in \mathcal{I}_k} \frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \exp\left[-\frac{1}{2} \bar{\mathbf{y}}_n^T \boldsymbol{\Sigma}_k^{-1} \bar{\mathbf{y}}_n\right] \end{aligned} \tag{H.7}$$

with $\boldsymbol{\Sigma}_k = s_k^2 \mathbf{K}_\mathbf{M} + \boldsymbol{\Sigma}_0$ ($k = 1, \ldots, K-1$), $\mathbf{s}^2 = [s_1^2, \ldots, s_K^2]^T$, $\mathbf{z} = [z_1, \ldots, z_N]^T$ and $\bar{\mathbf{y}}_n = \mathbf{y}_n - \mathbf{M}\boldsymbol{a}_n$. Moreover, $\mathcal{I}_k$ denotes the subset of pixel indexes belonging to the kth class and $\mathbf{K}_\mathbf{M} = \mathbf{Q}\mathbf{Q}^T$. The marginalized likelihood in (H.7) is identical to (4.18). Consequently, the proposed RCA-based model can be considered using (4.1) where the nonlinearity vectors $\boldsymbol{\phi}_n$ have been marginalized or using (H.1) where the nonlinearity parameters $\{\boldsymbol{\gamma}_n\}_{n=1,\ldots,N}$ have been marginalized. If the first interpretation has been preferred in the main body of Chapter 4 for its ease of understanding, the second interpretation provided in this Appendix is more rigorous in the sense that the densities involved in the derivation of the marginalized likelihood (H.7) are well defined. Conversely, the prior distributions in (4.11) can be degenerated since the rank of $\mathbf{K}_\mathbf{M}$ is equal to the rank of $\mathbf{Q}$ (i.e., $\mathrm{rank}(\mathbf{K}_\mathbf{M}) = R(R+1)/2$) which is often smaller than the number of spectral bands L.

Bibliography

Altmann, Y., Dobigeon, N., McLaughlin, S., and Tourneret, J. (2013a). Nonlinear spectral unmixing of hyperspectral images using Gaussian processes. *IEEE Trans. Signal Process.*, 61(10):2442–2453.

Altmann, Y., Dobigeon, N., McLaughlin, S., and Tourneret, J. (2013b). Residual component analysis of hyperspectral images - application to joint nonlinear unmixing and nonlinearity detection. *IEEE Trans. Image Processing.* submitted.

Altmann, Y., Dobigeon, N., and Tourneret, J. (2013c). Unsupervised post-nonlinear unmixing of hyperspectral images using a hamiltonian monte carlo algorithm. *IEEE Trans. Image Processing.* submitted.

Altmann, Y., Halimi, A., Dobigeon, N., and Tourneret, J. (2012). Supervised nonlinear spectral unmixing using a postnonlinear mixing model for hyperspectral imagery. *IEEE Trans. Image Processing*, 21(6):3017–3025.

Altmann, Y., Halimi, A., Dobigeon, N., and Tourneret, J.-Y. (2011). Nonlinearity detection in hyperspectral images using a polynomial post-nonlinear mixing model. Technical report, University of Toulouse, France.

Álvarez, M. A., Rosasco, L., and Lawrence, N. D. (2012). Kernels for vector-valued functions: A review. *Found. Trends Mach. Learn.*, 4(3):195–266.

Babaie-Zadeh, M., Jutten, C., and Nayebi, K. (2001). Separating convolutive post non-linear mixtures. In *Proc. of the 3rd Workshop on Independent Component Analysis and Signal Separation (ICA2001)*, pages 138–143, San Diego.

Bazaraa, M., Sherali, H., and Shetty, C. (1993). *Nonlinear programming: Theory and algorithms.* Wiley, 2 edition.

Berman, M., Kiiveri, H., Lagerstrom, R., Ernst, A., Dunne, R., and Huntington, J. F. (2004). ICE: A statistical approach to identifying endmembers in hyperspectral images. *IEEE Trans. Geosci. and Remote Sensing*, 42(10):2085–2095.

Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory.* John Wiley & Sons, New York.

Betancourt, M. J. (2010). Cruising The Simplex: Hamiltonian Monte Carlo and the Dirichlet Distribution. *ArXiv e-prints.*

Bioucas-Dias, J. M. and Nascimento, J. M. P. (2008). Hyperspectral subspace identification. *IEEE Trans. Geosci. and Remote Sensing*, 46(8):2435–2445.

Bioucas-Dias, J. M., Plaza, A., Dobigeon, N., Parente, M., Du, Q., Gader, P., and Chanussot, J. (2012). Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE J. Sel. Topics Appl. Earth Observations Remote Sensing*, 5(2):354–379.

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford University Press, Inc., New York, NY, USA.

Bonilla, E. V., Agakov, F. V., and Williams, C. K. I. (2007). Kernel multi-task learning using task-specific features. In *Proc. of the 11th Int. Conf. on Artificial Intelligence and Statistics (AISTATS)*.

Broadwater, J. and Banerjee, A. (2009). A comparison of kernel functions for intimate mixture models. In *in Proc. IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–4.

Broadwater, J., Chellappa, R., Banerjee, A., and Burlina, P. (2007). Kernel fully constrained least squares abundance estimates. In *Proc. IEEE Int. Conf. Geosci. and Remote Sensing (IGARSS)*, pages 4041–4044, Barcelona, Spain.

Brookes, M. (2005). *The Matrix Reference Manual*. Imperial College, London, UK.

Brooks, S. (2011). *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. Taylor & Francis.

Burazerovic, D., Heylen, R., Geens, B., Sterckx, S., and Scheunders, P. (2013). Detecting the adjacency effect in hyperspectral imagery with spectral unmixing techniques. *IEEE J. Sel. Topics in Applied Earth Observations and Remote Sensing,*, PP(99):1–9.

Chaudhry, F., Wu, C.-C., Liu, W., C.-I Chang, and Plaza, A. (2006). Pixel purity index-based algorithms for endmember extraction from hyperspectral imagery. In C.-I Chang, editor, *Recent Advances in Hyperspectral Signal and Image Processing*, chapter 2. Research Signpost, Trivandrum, Kerala, India.

Chen, J., Richard, C., and Honeine, P. (2013a). Estimating abundance fractions of materials in hyperspectral images by fitting a post-nonlinear mixing model. In *in Proc. IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–4, Gainesville, Florida.

Chen, J., Richard, C., and Honeine, P. (2013b). Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model. *IEEE Trans. Signal Process.*, 61(2):480–492.

Clark, R. N., Swayze, G. A., and Gallagher, A. (1993). Mapping minerals with imaging spectroscopy, U.S. Geological Survey. *Office of Mineral Resources Bulletin*, 2039:141–150.

Clark *et al.*, R. N. (2003). Imaging spectroscopy: Earth and planetary remote sensing with the USGS Tetracorder and expert systems. *J. Geophys. Res.*, 108(E12):5–1–5–44.

Close, R., Gader, P., Wilson, J., and Zare, A. (2012). Using physics-based macroscopic and microscopic mixture models for hyperspectral pixel unmixing. *Proc. SPIE Algorithms and Technologies ofr Multispectral, Hyperspectral and Ultraspectral pixel unmixing*, pages 83901L–83901L–13.

Craig, M. (1994). Minimum volume transforms for remotely sensed data. *IEEE Trans. Geosci. and Remote Sensing*, 32(3):542–552.

Dobigeon, N. and Févotte, C. (2013). Robust nonnegative matrix factorization for nonlinear unmixing of hyperspectral images. In *Proc. IEEE GRSS Workshop on Hyperspectral Image and SIgnal Processing: Evolution in Remote Sensing (WHISPERS)*, Gainesville, FL.

Dobigeon, N., Moussaoui, S., Coulon, M., Tourneret, J.-Y., and Hero, A. O. (2009a). Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery. *IEEE Trans. Signal Process.*, 57(11):2657–2669.

Dobigeon, N., Moussaoui, S., Tourneret, J.-Y., and Carteret, C. (2009b). Bayesian separation of spectral sources under non-negativity and full additivity constraints. *Signal Processing*, 89(12):2657–2669.

Dobigeon, N., Tourneret, J.-Y., and C.-I Chang (2008). Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery. *IEEE Trans. Signal Process.*, 56(7):2684–2695.

Eches, O., Dobigeon, N., Mailhes, C., and Tourneret, J.-Y. (2010). Bayesian estimation of linear mixtures using the normal compositional model. *IEEE Trans. Image Processing*, 19(6):1403–1413.

Eches, O., Dobigeon, N., and Tourneret, J.-Y. (2011). Enhancing hyperspectral image unmixing with spatial correlations. *IEEE Trans. Geosci. and Remote Sensing*, 49(11):4239–4247.

Fan, W., Hu, B., Miller, J., and Li, M. (2009). Comparative study between a new nonlinear model and common linear model for analysing laboratory simulated-forest hyperspectral data. *Remote Sensing of Environment*, 30(11):2951–2962.

Fauvel, M., Tarabalka, Y., Benediktsson, J., Chanussot, J., and Tilton, J. (2013). Advances in spectral -spatial classification of hyperspectral images. *Proceedings of the IEEE*, 101(3):652–675.

Gader, P., Dranishnikov, D., Zare, A., and Chanussot, J. (2012). A sparsity promoting bilinear unmixing model. In *in Proc. IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Shanghai, China.

Gorman, J. and Hero, A. (1990). Lower bounds for parametric estimation with constraints. *Information Theory, IEEE Transactions on*, 36(6):1285–1301.

Green, P. J. (1995). Reversible jump MCMC computation and Bayesian model determination. *Biometrika*, 82(4):711Ű732.

Halimi, A., Altmann, Y., Dobigeon, N., and Tourneret, J.-Y. (2011a). Nonlinear unmixing of hyperspectral images using a generalized bilinear model. *IEEE Trans. Geosci. and Remote Sensing*, 49(11):4153–4162.

Halimi, A., Altmann, Y., Dobigeon, N., and Tourneret, J.-Y. (2011b). Unmixing hyperspectral images using a generalized bilinear model. In *Proc. IEEE Int. Conf. Geosci. and Remote Sensing (IGARSS)*, pages 1886 –1889.

Han, T. and Goodenough, D. (2008). Investigation of nonlinearity in hyperspectral imagery using surrogate data methods. *Geoscience and Remote Sensing, IEEE Transactions on*, 46(10):2840–2847.

Hapke, B. W. (1981). Bidirectional reflectance spectroscopy. I. Theory. *J. Geophys. Res.*, 86:3039–3054.

Heinz, D. C. and C.-I Chang (2001). Fully constrained least-squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Trans. Geosci. and Remote Sensing*, 29(3):529–545.

Heylen, R., Burazerovic, D., and Scheunders, P. (2011). Non-linear spectral unmixing by geodesic simplex volume maximization. *IEEE J. of Sel. Topics in Signal Processing*, 5(3):534–542.

Heylen, R. and Scheunders, P. (2012). Calculation of geodesic distances in nonlinear mixing models: Application to the generalized bilinear model. *IEEE Geosci. and Remote Sensing Lett.*, 9(4):644–648.

Iordache, M. D., Bioucas-Dias, J. M., and Plaza, A. (2013). MUSIC-CSR: hyperspectral unmixing via multiple signal classification and collaborative sparse regression. *IEEE Trans. Geosci. and Remote Sensing*. submitted.

Jutten, C. and Karhunen, J. (2003). Advances in nonlinear blind source separation. In *4th Int. Symp. on Independent Component Analysis and Blind Signal Separation (ICA2003)*, pages 245–256, Nara, Japan.

Kalaitzis, A. A. and Lawrence, N. D. (2012). Residual components analysis. In *in Proc. ICML*.

Kay, S. M. (1993). *Fundamentals of Statistical Signal Processing: Estimation theory*. Prentice Hall, Englewood Cliffs NJ.

Kay, S. M. (1998). *Fundamentals of Statistical Signal Processing: Detection theory*. Prentice Hall.

Kent, J. T. and Mardia, K. (1988). Spatial classification using fuzzy membership models. *IEEE Trans. Patt. Anal. Mach. Intell.*, 10(5):659–671.

Keshava, N. and Mustard, J. F. (2002). Spectral unmixing. *IEEE Signal Processing Magazine*, pages 44–57.

Lawrence, N. D. (2003). Gaussian process latent variable models for visualisation of high dimensional data. In *NIPS*, Vancouver, Canada.

Lawrence, N. D. and Quiñonero Candela, J. (2006). Local distance preservation in the gp-lvm through back constraints. In *Proceedings of the 23rd international conference on Machine learning*, ICML '06, pages 513–520, New York, NY, USA. ACM.

Li, J. and Bioucas-Dias, J. M. (2008). Minimum volume simplex analysis: a fast algorithm to unmix hyperspectral data. In *Proc. IEEE Int. Conf. Geosci. and Remote Sensing (IGARSS)*, volume 3, pages 250–253, Boston, USA.

Licciardi, G., Ceamanos, X., Doute, S., and Chanussot, J. (2012). Unsupervised nonlinear spectral unmixing by means of NLPCA applied to hyperspectral imagery. In *Proc. IEEE Int. Conf. Geosci. and Remote Sensing (IGARSS)*, pages 1369–1372.

Mathews, V. J. and Sicuranza, G. L. (2000). *Polynomial Signal Processing*. Wiley, New York.

Meganem, I., Deliot, P., Briottet, X., Deville, Y., and Hosseini, S. (2013). Linear-quadratic mixing model for reflectances in urban environments. *IEEE Trans. Geosci. and Remote Sensing*, PP(99):1–15.

Miao, L. and Qi, H. (2007). Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization. *IEEE Trans. Geosci. and Remote Sensing*, 45(3):765–777.

Møller, M. F. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, 6(4):525–533.

Nascimento, J. M. and Bioucas-Dias, J. M. (2005). Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE Trans. Geosci. and Remote Sensing*, 43(4):898–910.

Nascimento, J. M. P. and Bioucas-Dias, J. M. (2009). Nonlinear mixture model for hyperspectral unmixing. In Bruzzone, L., Notarnicola, C., and Posa, F., editors, *Proc. SPIE Image and Signal Processing for Remote Sensing XV*, pages 7477–74770I–8. SPIE.

Nguyen, H., Richard, C., Honeine, P., and Theys, C. (2012). Hyperspectral image unmixing using manifold learning methods derivations and comparative tests. In *Proc. IEEE Int. Conf. Geosci. and Remote Sensing (IGARSS)*, pages 3086–3089.

Pakman, A. and Paninski, L. (2012). Exact Hamiltonian Monte Carlo for Truncated Multivariate Gaussians. *ArXiv e-prints*.

Papoulis, A. (1991). *Probability, random variables, and stochastic processes*. Communications and signal processing. McGraw-Hill, 2 edition.

Pereyra, M., Dobigeon, N., Batatia, H., and Tourneret, J.-Y. (2013). Estimating the granularity coefficient of a Potts-Markov random field within an MCMC algorithm. *IEEE Trans. Image Processing*, 22(6):2385–2397.

Perez-Cruz, F., Van Vaerenbergh, S., Murillo-Fuentes, J., Lazaro-Gredilla, M., and Santamaria, I. (2013). Gaussian processes for nonlinear signal processing: An overview of recent advances. *IEEE Signal Proc. Mag.*, 30(4):40–50.

Punskaya, E., Andrieu, C., Doucet, A., and Fitzgerald, W. (2002). Bayesian curve fitting using MCMC with applications to signal segmentation. *IEEE Trans. Signal Process.*, 50(3):747–758.

Quiñonero-candela, J., Rasmussen, C. E., and Herbrich, R. (2005). A unifying view of sparse approximate Gaussian process regression. *Journal of Machine Learning Research*, 6:2005.

Quirion, S., Duchesne, C., Laurendeau, D., and Marchand, M. (2008). Comparing GPLVM approaches for dimensionality reduction in character animation. *Journal of WSCG*, 16(1-3):41–48.

Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 31(2).

Rasmussen, C. E. and Williams, C. K. I. (2005). *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press.

Robert, C. P. (2007). *The Bayesian Choice: from Decision-Theoretic Motivations to Computational Implementation*. Springer Texts in Statistics. Springer-Verlag, New York, 2 edition.

Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer-Verlag, New York, second edition.

Robert, C. P. and Cellier, D. (1998). Convergence control of MCMC algorithms. In Robert, C. P., editor, *Discretization and MCMC Convergence Assessment*, pages 27–46. Springer Verlag, New York.

RSI (Research Systems Inc.) (2003). *ENVI User's guide Version 4.0.* Boulder, CO 80301 USA.

Scholkopf, B. and Smola, A. J. (2001). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond.* MIT Press, Cambridge, MA, USA.

Sheeren, D., Fauvel, M., Ladet, S., Jacquin, A., Bertoni, G., and Gibon, A. (2011). Mapping ash tree colonization in an agricultural mountain landscape: Investigating the potential of hyperspectral imagery. In *Proc. IEEE Int. Conf. Geosci. and Remote Sensing (IGARSS)*, pages 3672–3675.

Somers, B., Asner, G. P., Tits, L., and Coppin, P. (2011). Endmember variability in spectral mixture analysis: A review. *Remote Sensing of Environment*, 115(7):1603–1616.

Somers, B., Cools, K., Delalieux, S., Stuckens, J., der Zande, D. V., Verstraeten, W. W., and Coppin, P. (2009). Nonlinear hyperspectral mixture analysis for tree cover estimates in orchards. *Remote Sensing of Environment*, 113(6):1183–1193.

Tarabalka, Y., Fauvel, M., Chanussot, J., and Benediktsson, J. (2010). Svm- and mrf-based method for accurate classification of hyperspectral images. *IEEE Geosci. and Remote Sensing Lett.*, 7(4):736–740.

Urtasun, R., Fleet, D. J., and Lawrence, N. D. (2007). Modeling human locomotion with topologically constrained latent variable models. In *Conf. Human motion: understanding, modeling, capture and animation*, pages 104–118, Rio de Janeiro, Brazil. Springer-Verlag.

Wang, J. and C.-I Chang (2006). Applications of independent component analysis in endmember extraction and abundance quantification for hyperspectral imagery. *IEEE Trans. Geosci. and Remote Sensing*, 4(9):2601–2616.

Winter, M. (1999). Fast autonomous spectral end-member determination in hyperspectral data. In *Proc. 13th Int. Conf. on Applied Geologic Remote Sensing*, volume 2, pages 337–344, Vancouver.

Yokoya, N., Chanussot, J., and Iwasaki, A. (2012). Generalized bilinear model based nonlinear unmixing using semi-nonnegative matrix factorization. In *Proc. IEEE Int. Conf. Geosci. and Remote Sensing (IGARSS)*, pages 1365–1368.

Zare, A. and Gader, P. (2007). Sparsity promoting iterated constrained endmember detection in hyperspectral imagery. *IEEE Geoscience and Remote Sensing Letters*, 4(3):446–450.

Zare, A., Gader, P., Allgire, T., Drashnikov, D., and Close, R. (2012). Bootstrapping for piece-wise convex endmember distribution detection. In *Proc. IEEE GRSS Workshop on Hyperspectral Image and SIgnal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–4.